TABLE OF CONTENTS

SYNTHETIC VISION SYSTEMS: FLIGHTPATH TRACKING, SITUATION AWARENESS, AND VISUAL SCANNING IN AN INTEGRATED HAZARD DISPLAY

Amy L. Alexander and **Christopher D. Wickens**
University of Illinois, Aviation Human Factors Division
Savoy, Illinois

Twenty-four certified flight instructors participated in an experiment designed to examine the viability of three Integrated Hazard Display (IHD) formats representative of Synthetic Vision System (SVS) technology (2D coplanar, 3D exocentric, split-screen; Wickens, 2003) in supporting flightpath tracking and situation awareness (SA). SA was probed through the use of two techniques, a memory-based technique called SAGAT and a variant of a perception-based technique called SPAM. Overall, the 3D exocentric display appeared to be the worst display format in terms of supporting SA and utilizing visual attention for the betterment of performance. There was an apparent speed-accuracy tradeoff between the memory-based (display blank) and perception-based (display present) conditions such that pilots took longer to make their traffic position estimations when the display was present, but those judgments were made with greater accuracy compared to when the display was removed. The perception-based measurement technique appeared to be the most sensitive to display differences in supporting SA.

Introduction

Synthetic vision systems (SVS) have been proposed as a possible solution to such problems in aviation as controlled flight into terrain and low-visibility conditions (Alexander, Wickens, & Hardy, accepted; Prinzel, Comstock, Glaab, Kramer, & Arthur, 2004; Schnell, Kwon, Merchant, & Etherington, 2004). SVS provides an artificial, real-time presentation of terrain and traffic to enhance situation awareness (SA), combined with a depiction of the planned trajectory from a 3D perspective to support guidance and control (Williams, Waller, Koelling, Burdette, Doyle, Capron, Barry, & Gifford, 2001).

While a primary flight display (PFD) has been developed to provide tunnel flightpath guidance, it may or may not also be used to represent other hazards such as terrain or traffic aircraft. In the absence of such information within the PFD itself, a critical component of the SVS suite becomes the Integrated Hazard Display (IHD). IHDs are specifically being developed to assist in navigational tasks by representing terrain and traffic hazards through the use of high-resolution terrain databases and satellite-based navigation systems. However, the best perspective from which to present IHD information is still under investigation as research has generally offered conflicting results as to which of many display options are most optimal for the various tasks involved with navigation. The goal of the current study is to examine flight performance, situation awareness (SA), and visual scanning in the context of three IHD frame of reference formats: the 2D coplanar, 3D exocentric, and split-screen displays.

A 2D coplanar display contains a top-down view of the flight environment in the top panel, as well as a side-view depiction in the bottom panel, also called a vertical situation display (VSD; Fadden, Braune, & Wiedemann, 1993; Thearle, 2002). More precise spatial and relative position judgments are best made using a 2D coplanar display due to its unambiguous depiction of the three dimensional airspace (St. John, Smallman, Bank, & Cowen, 2001; Wickens, 2000). Despite its faithful axis representation, the 2D coplanar display imposes a visual scanning cost due to the presentation of lateral and vertical information on two different display panels. This spatial separation of information will produce information access costs (IACs) to the extent that cognitive and/or physical effort must be exerted in sampling the two views (Wickens, 1992).

While 3D displays have been supported due to their "natural", integrated representation of the 3D world, costs in terms of biases and distortions are inherent. Namely, the "2D-3D effect" leads pilots to subjectively rotate vectors in depth more parallel to the viewing plane (McGreevy & Ellis, 1986). This effect may be manifest as the compression effect which describes how at least two of three axes must be compressed to display a 3D world on a 2D screen. Increased compression is associated with a reduction in resolution which will lead to a bias in estimating distances along the compressed axis as shorter than they really are (Boeckman & Wickens, 2001).

One possible solution to the tradeoffs between 2D and 3D displays is the "split-screen display", consisting of a 3D exocentric view to support global awareness and the side-view VSD of the 2D coplanar format to support precise hazard localization and avoidance. Although split-screen displays resolve issues of bias and distortion associated with 3D displays by also providing a VSD which inherently maintains faithful axis representations, inappropriate allocation of visual attention to the more compelling and information-rich 3D exocentric panel may deter performance overall, as found in previous work involving a split-screen display (Olmos, Wickens, & Chudy, 2000; note that this study used 3D exocentric and 3D egocentric panels, without a VSD).

Given the importance of SA maintenance in preventing incidents from occurring under low-visibility or terrain-challenging conditions, we now turn to the issue of measuring SA. SA can be defined as “the perception of the elements within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” (Endsley, 1995, p. 36). Endsley (1988) has proposed a memory-based Situation Awareness Global Assessment Technique (SAGAT), in which a scenario is temporarily frozen and hidden from view while the pilot is asked a series of questions concerning the location of entities within the display. These questions must be answered by consulting working memory or long-term working memory.

It has been argued that having high situation awareness does not necessarily require **memory** of relevant information. Durso and colleagues (1998) proposed that knowing where to find information could be indicative of good situation awareness even if that information was not available in memory. In light of this, a Situation Present Assessment Methodology (SPAM) was developed which would rely on perception of the situation at hand in answering real-time probes.

Analysis of these techniques suggests the likelihood of a speed-accuracy tradeoff. SA measures of perception (e.g., SPAM) may lead to greater accuracy, given that the original data are available for inspection, but this would be at the cost of a longer response time since it will take time to process that information. These results, of course, would be relative to lower accuracy and faster response times with SA measures of memory (e.g., SAGAT) given that without the original data available pilots will be forced to rely upon a degrading memory trace. Such a tradeoff was indeed found in a previous study examining traffic awareness within an IHD context (Alexander & Wickens, 2004).

The current paper describes results from a study which examined flightpath tracking, SA, and visual scanning to assess attention allocation within an IHD context. A PFD containing a tunnel-in-the-sky was presented in the upper-left corner of the screen, while the IHD was presented to the right of it. Given that the PFD provided tunnel guidance, the format of which was consistent across IHD presentations, we do not expect to see differences in flightpath tracking across display types. Any differences therein, however, would presumably be governed by the extent to which the IHD demanded attention from the pilot, a quantity inferred in the present experiment from the measure of visual scanning.

SA, or more specifically, traffic awareness, was probed through SAGAT and SPAM. Our SAGAT probes consisted of freezing the simulation and blanking the IHD at unexpected times and asking pilots to estimate the position of a queried aircraft in the outside world based on its representation within the IHD (note that aircraft were not visible in the outside world). Our SPAM-variant also consisted of freezing the simulation, although the IHD and queried traffic remained visible. SA as measured by traffic probes will presumably be better supported by a 2D coplanar or split-screen display than a 3D exocentric display due to the faithful axis representation within the former formats (both panels of the 2D coplanar, bottom panel of the split-screen).

The display modulation of flightpath tracking and SA traffic position estimation performance will also be examined in terms of visual scanning measures of pilot attention allocation. Such measures are hypothesized to reveal (dis)associations with performance to the extent that relations of changing performance and/or scanning behavior across conditions can speak to the nature of the underlying processes. For example, in terms of flightpath tracking performance, equivalent performance is predicted across display types. Scanning measures might reveal, however, that less visual attention is demanded in a specific display, therefore allowing more visual attention to be freed for other tasks. The freeing of visual resources may be seen as an advantage to that display despite equivalent flightpath tracking performance, given that the flight environment is often composed of multiple task demand at any given time.

Method

Twenty-four certified flight instructors (age, *M* = 21.6; experience, *M* = 514 total flight hours, *M* = 83 instrument flight hours) from the University of Illinois Institute of Aviation flew a series of flightpaths and made judgments regarding traffic locations based on the representations of three IHD formats. The experiment was conducted on a high-fidelity Frasca flight simulator with a 180° outside-world view spread across three display screens. Pilots were paid $9/hour for their participation.

Displays

2D Coplanar. The coplanar display shown in Figure 1a consisted of two windows offering a horizontal, top-down (X-Z axes) view and a vertical, side-looking (Y-Z axes) VSD projected orthogonally (without perspective information) depicting 4 miles ahead of ownship and 1 mile behind. The terrain in the top-down panel is color-coded relative to ownship: red represents terrain that is higher than ownship, yellow represents terrain that is up to 1000ft lower than ownship, black represents terrain that is more than 1000ft lower than ownship. A predictor vector based on current state information was displayed.

3D Excocentric. The 3D exocentric display presented a "tethered" view (see Figure 1b). An elevation angle of 45° was imposed to optimize judgments within the longitudinal and vertical dimensions (Boeckman & Wickens, 2001) with an azimuth offset of approximately 10° in the clockwise direction (Ellis, McGreevy, & Hitchcock, 1987). The ambiguity of judgments in the vertical direction was further reduced by attaching a "drop line" from ownship and other aircraft to the terrain below (St. John, Cowan, Smallman, & Oonck, 2001: Wickens, 2003). A predictor vector based on current state information was displayed.

Split-Screen. The split-screen view was comprised of a 3D exocentric view in the top panel and a side-view VSD in the bottom panel (see Figure 1c). A predictor vector based on current state information was displayed.

(a) (b) (c)

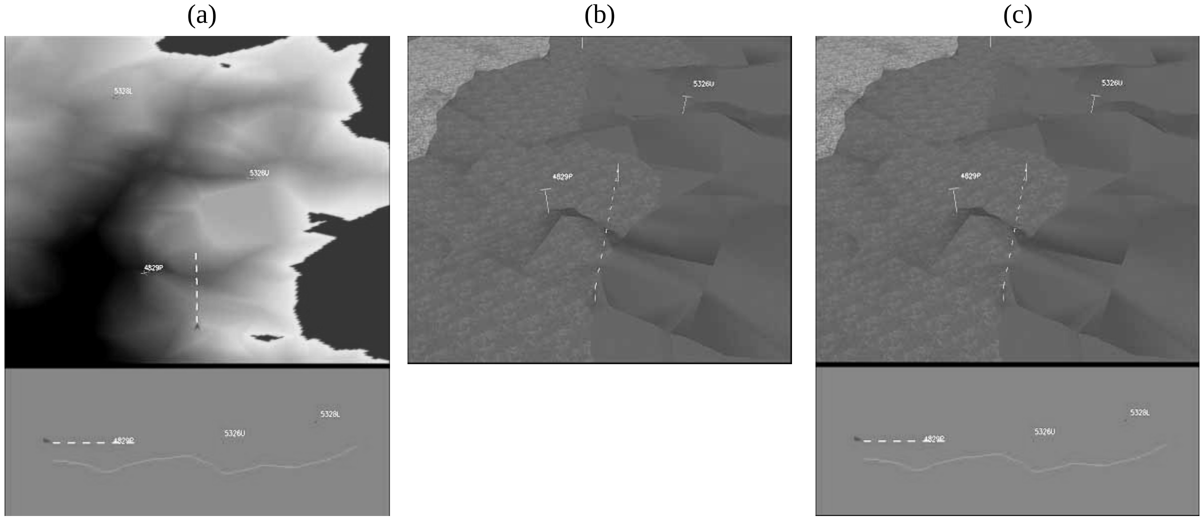


Figure 1. *Display formats: (a) 2D coplanar display, (b) 3D exocentric display, and (c) split-screen view.*

Task & Design

Pilots made traffic location judgments on a total of 60 aircraft targets across the three IHD types. Pilots flew scenarios containing multiple aircraft, between one and four of which were within the display view at any given time. Pilots were periodically asked, during simulation freezes, to estimate the location of the nearest aircraft within the outside world. Visibility was adjusted so that these aircraft were <u>not</u> visible in the outside world. However, the outside world did present the corresponding mountainous terrain that was visible on the display, so that correspondence between locations in the outside world and the display could be easily established. During simulation freezes on some trials the display would remain visible (SPAM-variant), whereas on others, it would blank (SAGAT).

Upon one of these two events occurring, the pilot was first asked to use a knob on the left-hand of the yoke to move a white ball in the outside world to the position where they estimated the location of the closest aircraft to be. Once the pilot placed the white ball in the desired location, s/he pressed a button on the yoke to continue the scenario. Pilots were instructed to perform the location estimation task as quickly and accurately as possible.

A within-subjects manipulation of IHD format was used. The presentation of IHD format was counterbalanced so that every possible combinatory order of the formats was used, and then repeated in reverse order. Display presentation was counterbalanced across pilots. The two display present/blank conditions described previously were quasi-randomized within each scenario.

Eye movements were recorded by an Applied Systems Laboratory (ASL) Model 5000 eye-tracker throughout the experiment. Those data collected during the simulation freezes were removed from analysis.

Results

Flightpath Tracking Performance. Given that flightpath information was presented identically across display types (that is, shown in the egocentric PFD in the upper-left corner of the display), it is not surprising that there were no main effects of display type in either vertical or lateral deviations ($F(2, 46) = 0.87$, $p > .42$; $F(2, 46) = 1.12$, $p > .33$, respectively).

SA Response Time. Results revealed a significant main effect of SA measurement condition ($F(1, 23) = 43.2$, $p < .001$) such that response time to the traffic awareness probes was two seconds faster in the memory (display blank; $M = 6.44$ s) than perception (display present; $M = 8.44$ s) condition. There was no effect of display nor an interaction of display type and condition (both $p > .24$).

Vertical Position Estimation Error. As shown in Figure 2, vertical estimation error results revealed a significant main effect of SA measurement condition ($F(1, 23) = 33.6$, $p < .001$) such that estimation error was about two degrees of visual angle greater in the memory (display blank; $M = 6.64$ degrees) than perception (display present; $M = 5.15$ degrees) condition.

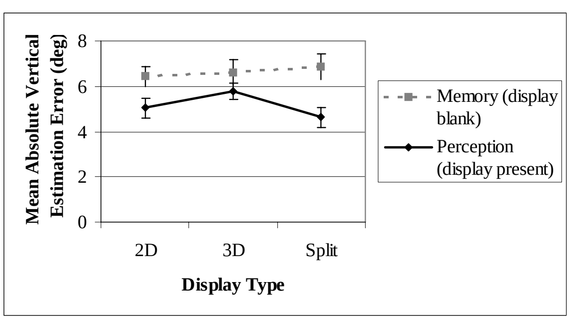


Figure 2. *Mean absolute vertical estimation error by display type and condition.*

Although there was no effect of display nor an interaction of display type and condition (both $p > .22$), there was a significant difference within the perception (display blank) limb such that vertical estimation error was about 1.5 degrees greater with the 3D ($M = 5.78$ degrees) than split-screen ($M = 4.63$ degrees) display ($t(23) = 2.53$, $p < .02$).

Lateral Position Estimation Error. There was a significant main effect of condition ($F(1, 23) = 25.4$, $p < .001$) such that lateral estimation error was about four degrees of visual angle greater in the memory (display blank; $M = 11.9$ degrees) than perception (display present; $M = 7.93$ degrees) condition. There was no effect of display nor an interaction of display type and condition (both $p > .26$).

Mean Percent Dwell Time. The allocation of attention, as measured by percent dwell time (PDT) within the different areas of interest (AOI), is shown in Figure 3. Again, these measures do not reflect scanning during simulation freezes. Results reveal an obvious dominance of scans to the PFD about 66% of the time in all display conditions. Visual attention was captured roughly 19% of the time by the top panel of the IHD, regardless of whether that panel presented a 2D or 3D view. Scanning to the VSD and outside world was equivalent between the 2D coplanar and split-screen formats, accounting for about 8% of the time, within the 2D coplanar and split-screen displays. Given that the 3D exocentric display format did not have a VSD representation, the extra visual attention which had been directed to the VSD in the other two displays was instead split among the PFD and top panel of the IHD (i.e., the 3D view).

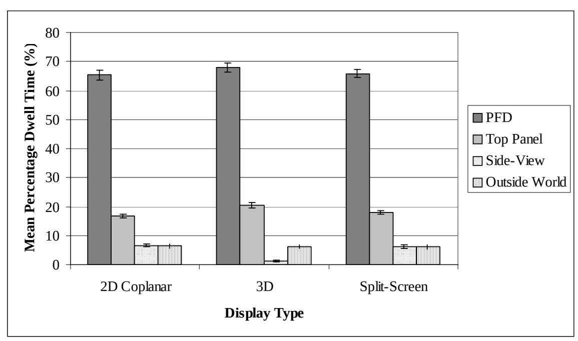


Figure 3. *Mean percentage dwell time by display type and area of interest.*

In terms of effects driven by the attentional demands of the IHD formats, a few differences within the individual AOIs are of interest. First, visual attention was directed to the PFD about 2% of the time more with the 3D exocentric display than either the 2D coplanar or split-screen views ($F(2, 46) = 3.27$, $p < .05$). Pilots also spent about 3% more time looking at the IHD with the 3D exocentric display compared to visual scans to the 3D panel of the IHD in the 2D coplanar and split-screen views ($F(2, 46) = 16.7$, $p < .001$). However, pilots spent less time looking at the IHD in the 3D display than they spent looking at both panels of the IHD in the other two formats.

Discussion

In examining the null effects of display type within the flightpath tracking data, it is apparent that pilots were protecting the primary flight task of aviating and navigating. In other words, attentional demands of the different IHD formats did not affect tracking performance as pilots were appropriately treating that task as top priority.

Added visual attention to the PFD in the 3D exocentric condition did not improve flightpath tracking performance relative to that obtained with the 2D coplanar and split-screen displays. Increased scans to the IHD with the 3D exocentric compared to the 2D coplanar and split-screen displays also showed no improvement in terms of estimating traffic position during the SA probes, and indeed, position estimation error within the vertical dimension, in fact, was worst with the 3D display (in the display present condition). SA within the 3D exocentric display was expected to be more poorly supported due to the lack of a faithful presentation of the vertical dimension. Hence, the added visual attention to the IHD was not enough to resolve the ambiguities inherent to a 3D exocentric viewpoint.

Interestingly, the only display difference found in the SA data was revealed within the perception-based (display present) SPAM condition. As already discussed, traffic position estimation was found to be better supported by the split-screen than 3D display when examining judgments specifically within the vertical dimension. This finding of the SPAM condition being most sensitive to display differences requires further exploration.

In terms of the specific traffic awareness measures used in this study, there was an apparent speed-accuracy tradeoff between the memory-based (display blank) and perception-based (display present) conditions. While pilots took longer to make their traffic positions estimation when the display was present, those judgments were made with greater accuracy compared to when the display was removed. As described in the introduction, such a tradeoff was expected given that more perceptual data was available during display-present SPAM simulation freezes, and it therefore took pilots longer to process the available information. The consequence of this longer processing, however, is for improved accuracy relative to the degraded memory trace available in the display-blank SAGAT freezes.

Conclusions

This study not only examines dimensionality within an important context for aviation safety (an SVS IHD), it also addresses a relatively new design concept which brings the "best (or worst) of both worlds" (i.e., 2D coplanar and 3D displays) together in a split-screen format. Importantly, the 3D exocentric display appeared to be the worst display format in terms of supporting SA and utilizing visual attention for the betterment of performance. Thus highlighting the critical importance of a VSD for hazard awareness (Fadden et al., 1993; Thearle, 2002). Importantly, while such a VSD "consumes" slightly more attentional resources to process, the withdrawal of these resources from the PFD led to no decline in performance.

Equally important is the comparison of SA methodologies within a traffic awareness framework. A speed-accuracy tradeoff is noted between the perception-based (SPAM) and memory-based (SAGAT) conditions such that pilots took longer to make their traffic position estimations when the display was present, but those judgments were made with greater accuracy compared to when the display was removed. The perception-based measurement technique appeared to be the most sensitive to display differences in supporting SA task, although display differences were only found within the vertical dimension position estimations.

These flightpath tracking, SA, and visual scanning findings have implications for both the design of an IHD in terms of display format, and the evaluations which lead to the recommendations therein. Given the overall lack of display differences found, specifically between the 2D coplanar and split-screen views, more experimentation is recommended in resolving what types of tasks one format might be better than the other in supporting. We have only examined one task in the current study, traffic awareness, one of a general class of SA measures. More comprehensive conclusions with regard to global awareness, hazard localization, and hazard avoidance measures are desired in recommending a single IHD format.

Acknowledgments

This research was supported by contract NASA-NAG-1-03014 from the NASA Langley Research Center, for which Dr. Lance Prinzel was the technical monitor. The authors wish to thank Ron Carbonari for his software development and post-processing contributions, and T.J. Hardy for his help with scenario development.

References

Alexander, A.L. & Wickens, C.D. (2004). Measuring traffic awareness in an integrated hazard display. In *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: HFES.

Alexander, A.L., Wickens, C.D., & Hardy, T.J. (accepted). Synthetic vision systems: The effects of guidance symbology, display size, and field of view. *Human Factors*.

Boeckman, K.J. & Wickens, C.D. (2001). *The resolution and performance effects of three-dimensional display rotation on local guidance and spatial awareness measures* (Technical Report ARL-01-4/NASA-01-3). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Fadden, D.M., Braune, R., & Wiedemann, J. (1993). Spatial displays as a means to increase pilot situational awareness. In Ellis, S.R., M. Kaiser, & A.J. Grunwald (Ed.), *Pictorial Communication in Virtual and Real Environments* (pp. 172-181). London: Taylor & Francis.

McGreevy, M.W. & Ellis, S.R. (1986). The effect of perspective geometry on judged direction in spatial information instruments. *Human Factors, 28*.

Prinzel, L.J., Comstock, J.R., Glaab, L.J., Kramer, L.J., & Arthur, J.J. (2004). The efficacy of head-down and head-up synthetic vision display concepts for retro- and forward-fit of commercial aircraft. *International Journal of Aviation Psychology, 14*.

Schnell, T., Kwon, Y., Merchant, S., & Etherington, T. (2004). Improved flight technical performance in flight decks equipped with synthetic vision information system displays. *International Journal of Aviation Psychology, 14*.

Scott, W.B. (2001, October 29). Synthetic vision systems: Ready for prime time? *Aviation Week & Space Technology*, 78-80.

Smith, D.R. (2001). *Controlling pilot error: Controlled flight into terrain (CFIT/CFTT).* McGraw-Hill Professional Publishing.

St. John, M., Cowen, M.B., Smallman, H.S., & Oonk, H.M. (2001). The use of 2D and 3D displays for shape-understanding versus relative-position tasks. *Human Factors, 43* (1), 79-98.

Thearle, C. (2002, March 19). Boeing introduces the technology demonstrator airplane: Boeing commercial airplanes gives flight to new, emerging technologies (Boeing News Release). Seattle, WA: Boeing Commercial Airplanes.

Wickens, C.D. (2000). The when and how of using 2-D and 3-D displays for operational tasks. In *Proceedings of the IEA 2000/HFES 2000 Congress*. Santa Monica, CA: HFES.

Wiener, E.L. (1977). Controlled flight into terrain accidents: System induced errors. *Human Factors, 19*.

Williams, D., Waller, M., Koelling, J., Burdette, D., Doyle, T., Capron, W., Barry, J., & Gifford, R. (2001). *Concept of operations for commercial and business aircraft synthetic vision systems* (Version 1.0, Draft Technical Memorandum). Hampton, VA: NASA LARC.

COGNITIVE SYSTEMS ENGINEERING APPROACH TO SHARED SITUATION AWARENESS FOR UNMANNED AERIAL VEHICLES

Matthijs Amelink
Decis lab
Delft, the Netherlands

Max Mulder & **M.M. (René) van Paassen**
Delft University of Technology, Faculty of Aerospace Engineering, Control and Simulation Division
Delft, the Netherlands

John Flach
Wright State University, department of Psychology
Dayton, Ohio

Integration of UAVs with Air Traffic Control (ATC) is a world wide problem. ATC is already troubled by capacity problems due to a vast amount of air traffic. In the future when large numbers of Unmanned Aerial Vehicles (UAVs) will participate in the same airspace, the situation cannot afford to have UAVs that need special attention. Regulations for UAV flights in civil airspace are still being developed but it is expected that authorities will require UAVs to operate "like manned aircraft". The implication is that UAVs need to become full participants of a complex socio-technical environment and need to generate 'man like' decisions and behavior. In order to deal with the complexity a novel approach to developing UAV autonomy is needed, aimed to create an environment that fosters shared situation awareness between the UAVs, pilots and controllers. The underlying principle is to develop an understanding of the work domain that can be shared between people and UAVs. A powerful framework to represent the meaningful structure of the environment is Rasmussen's abstraction hierarchy. This paper proposes that autonomous UAVs can base their reasoning, decisions and actions on the abstraction hierarchy framework and communicate about their goals and intentions with human operators. It is hypothesized that the properties of the framework can create 'shared situation awareness' between the artificial and human operators despite the differences in their internal workings.

Introduction

There now seems little doubt that UAVs will be part of civil aviation's future infrastructure. UAVs are Unmanned Aerial vehicles that are either remotely controlled from a base station or are autonomous. Ground control of the UAV varies from stick and rudder control, to performing navigation task to mission execution by the press of a button.

The military has been using UAVs for a variety of purposes. Their missions have been characterized as the "dull, dangerous and dirty" – missions that human pilots would typically not want to fly or are not suitable to fly. There are also plenty of these missions on the commercial civil side and include environmental and geological surveys, weather reporting, search and rescue, forest-fire monitoring, border patrol and communications relaying. (Reynish, 2004)

Most UAVs are designed to fly their mission below 40,000 feet in controlled airspace, which is airspace already heavily populated by manned aircraft. In order to carry out these missions UAVs must be able to fly among conventional air traffic without demanding special handling by ATC. This would have an unacceptable impact on ATC workload and airspace capacity.

Although military UAV markets have been steadily growing, civil UAV applications have been slow to take advantage of potential applications. The slow start is, at least partially, due to the lack of a regulatory framework. Existing regulations cannot accommodate civil UAVs. A regulatory framework is being developed to ensure safety of UAV operations and allow seamless integration in national and international airspace. There are various initiatives world wide that aim to develop regulations; often they are partnerships between government and industry. Two examples are UVS-International initiated in Europe and Access Five in the United States. Despite the current lack of regulations, it is expected that regulatory authorities worldwide will require UAVs to operate identically to manned aircraft in civil controlled airspace (Avionics Magazine, October 2004). This is a major challenge UAV system design.

Concept UAV Regulations

Europe's UAV Task Force is a joint JAA / EUROCONTROL initiative to commence work leading to European regulations for civil UAVs. In May 2004 the UAV Task Force delivered: *A concept for European regulations for civil unmanned aerial vehicles*. In this report three of the guidelines that have been established during the development of the regulation stand out with respect to this research. They are repeated here shortly:

<u>Fairness</u>*: Any regulatory system must provide fair, consistent and equitable treatment of all those it seeks to regulate.*

<u>Equivalence</u>*: Regulatory standards should be set to be no less demanding than those currently applied to comparable manned aircraft nor should they penalize UAV Systems by requiring compliance with higher standards simply because technology permits. UAV operations shall not increase the risk to other airspace users or third parties. UAV operators should seek to operate within existing arrangements.*

<u>Transparency</u>*: The provisions of an Air Traffic Service (ARS) to a UAV must be transparent to the Air Traffic Control (ATC) controller and other airspace users. (…) UAVs must be able to comply with ATC instructions and with the equipment requirements applicable to the class of airspace within which they intend to operate.*

The U.S. Department of Defense faces the problem of enabling their military UAVs to fly in civil airspace. The Office of the Secretary of Defense has provided the *Airspace Integration Plan for Unmanned aviation* (2004). Two of the principles guiding their approach are repeated here:

<u>Do no harm</u>*: avoid new initiatives that would adversely impact air traffic control procedures and manned aviation.*

<u>Conform rather than create</u>*: avoid the creation of dedicated UAV regulations as much as possible. The goal is to achieve transparent flight operations in the National Airspace System.*

These guidelines and principles will have a great impact on how future UAV Systems will be designed to comply with the regulations internationally. For the most part they indicate that UAVs should fit in seamlessly with manned aviation and meet equivalent levels of safety. Only time will tell, when the actual regulations are enacted, how much room is left for dedicated UAV regulations.

All UAVs have to meet the regulations whether the UAV is autonomous or piloted from the ground. Those UAVs that depend on a communication link for control are sensitive to failure of that link. Failure may be due to e.g., atmospheric disturbances, hijacking attempts, jamming or tactical maneuvering. In any case in civil airspace, the UAV must ensure its safety and that of the other airspace users. How regulation will precisely deal with this mode of failure is unclear but it has been suggested that every UAV will need an autonomous mode that is capable of sense and avoid to ensure safety (Airspace Integration plan, 2004; UAV Task Force 2004). In the next paragraph the problems associated with developing UAV autonomy are addressed.

Another obstacle, and technological challenge, is that present UAVs cannot yet detect manned aircraft and conflict situations. Therefore they cannot safely share airspace with manned aircraft. To become accepted in civil airspace, UAVs need to have the capability to 'sense and avoid' other aircraft in their operating environment with the same level of safety as human pilots. This problem will also be addressed in the next paragraphs.

Problem Formulation

In the air traffic domain rules, procedures and regulations have centered on the way humans communicate and on human cognitive capabilities. The focus of the problem is on how human operators communicate about the meaning in the domain and build their situation awareness. For autonomous UAVs to effectively behave like manned aircraft, they need to be able to communicate about the same meaning and therefore share the same kind of situation awareness with human operators.

There are three areas of interest with respect to UAV behavior. To be a full participant in the airspace a UAV must be:

1. capable to sense and avoid other aircraft and obstacles.
2. a full participant in the ATC environment
3. able to cope with unanticipated events

1. Sense and avoid A lot of emphasis is put on 'sense and avoid' capability in the conceptual regulations because it is an important capability of critical safety concern. To us the term 'sense and avoid' seems incomplete because it omits the decision process that intermediates 'sense' and 'avoid'. Assuming that obstacles and other aircraft can be sensed, the weight of the problem is in deciding what action to take. Part of this process is assessing the situation and possibly negotiating a solution. Situation awareness plays an important role in this and the 'sense and avoid' capability is therefore seen as an integrated part of the overall autonomy and decision making architecture of the UAV.

2. Full participation in the ATC environment Controlled airspace is a complex socio-technical environment that is shared by many people that contribute to the system and are interdependent for function and safety. UAVs should become part of this environment and therefore integrating UAVs with existing ATC is not a matter of programming the optimal solutions for the problem but instead it is a matter of finding best human practice. A purely technological solution alone will not address the full scope of the integration problem; hence human factors is a core element of the UAV integration process.

Communication is the most important interface between the UAV and the ATC environment. It allows parties to share information, express intentions and resolve conflicts. It is unlikely that UAVs will communicate through speech but it will need to be able to use the concepts used in ATC and understand their meaning. How a UAV can understand meaning is related to how it can have situation awareness.

3. Unanticipated events This topic is left untouched by the concept regulations. It is the area where CSE is thought to have its major contribution. In the Airspace Integration Plan (2004) for unmanned aviation it is suggested that "Preprogrammed decision trees are built to address each possible failure during each part of the mission" (airspace Integration plan for unmanned aviation, office of the secretary of defense, 2004). Although this technique will cover a lot of failure modes in possibly a very effective way, there will always be some failures that were not anticipated by the designers. To ensure an equal level of safety as manned flight, UAVs need to be able to effectively cope with unanticipated events. To improvise and come up with new solutions to new problems requires an understanding of the structure of the work domain. The UAV needs to have this understanding / awareness.

A Domain Representation for UAVs

The difficult question is: "how to create machine situation awareness that is compatible with human situation awareness?" The answer lays in how the domain is represented internally: the UAV's mental model of the work domain has to be compatible with how human operators think of the work domain. The internal model will also determine the UAV's capabilities of dealing with the environment. We believe that part of the solution is in how people make abstractions in their work domain and that the properties of Rasmussen's abstraction hierarchy (Rasmussen, Peijtersen, and Goodstein, 1994) are central to this approach. To satisfy the requirements pointed out earlier, the abstraction hierarchy is proposed as the basis for the domain representation for autonomous UAVs.

The abstraction hierarchy is proposed as the basis for a domain representation mainly because its properties that are important to work domain analysis are also important for the intended domain representation. As described by Vicente (1999), the first important property is the psychologically useful way it represents complex work domains. The second important property is that it provides an informational basis for coping with unanticipated events. Both are shortly discussed below.

Psychological relevance The abstraction hierarchy consists of multiple domain representations on different levels of abstractions that are linked through functional means-ends relations. This type of hierarchy is explicitly purpose oriented and allows operators to deal with complexity effectively. Each level describes the domain but moving up the levels there is less detail and more purpose and meaning. Thus the top level describes the domain's functional goals which are usually abstract and the lowest level describes the physical implementation. For the air traffic domain you will find abstract terms like traffic flows, safety, and efficiency in the upper levels and more concrete terms like flight path, aircraft, and engine in the lower levels. Note that the abstraction hierarchy intended here covers the air traffic domain and not only the UAV system.

The abstraction hierarchy connects the elements of the work domain in means-ends manner so that they can be seen in relation to what their meaning is. This is the property that allows goal oriented problem solving. The problem solving itself is constrained to that which is relevant by starting on a high level of abstraction, moving down only concentrating on the subset of the domain that is connected to the function of interest. This allows for computationally economic problem solving (Vicente and Rasmussen, 1992) which is important to all resources limited agents.

To be transparent UAV decisions and actions should be based on a domain representation similar to that of the human operators. A domain representation is needed that is compatible with human thinking. The psychological relevance of the abstraction hierarchy has this implication. If the abstraction hierarchy is indeed psychologically relevant and people do reason within an abstraction hierarchy representation, it can form the common language in a socio-technical system. In other words; when a domain representation that is based on the abstraction hierarchy is successfully implemented in a UAV it should be able to deal with the domain complexity in a goal oriented way and communicate about the meaningful concepts in the domain. It should generate behavior that is compatible with the human way of dealing with the same problems. This is a first step towards man-like behavior as will be required by authorities.

Coping with unanticipated events Unanticipated events are by definition not foreseen by designers. Currently systems are not very good at dealing with these events and they form a big threat to safety. In ecological interface design (EID) the abstraction hierarchy representation provides a basis for coping with unanticipated events (Vicente and Rasmussen 1992). The abstraction hierarchy framework is used because it captures the domain complexity while it does not have built-in rules or procedures for dealing with the complexity. The work domain is described in terms of constraints that it imposes on the operator and does not describe actions or tasks to deal with the domain. When constraints are broken or not met, which will happen when the actual behavior and intended behavior differ, the representation provides a framework for goal directed problem solving.

This is in contrast to programming decision trees that address each possible failure during each part of the mission, but leave the unanticipated events unaccounted for. The abstraction hierarchy is constrained based and not rule based thus attention needs to shift from rule based reasoning to constraint based reasoning. The idea is that in combination with constraint based programming the representation can be used to deal with situations that wouldn't be captured in a rule-based knowledge system. It can be used to cope with unanticipated events.

As with EID, the abstraction hierarchy is used to support knowledge based behavior. However, it is not intended to engage in problem solving activity for every encountered situation. Rule based and skill based behavior can be much more computationally effective to apply to known solutions. To make this distinction the system will need to detect whether a known solution will be effective or if it needs to generate a new solution in a new situation. It is hypothesized that the abstraction hierarchy representation can support making this distinction.

Conclusion

The main benefit of developing this architecture is the psychological relevance it has. It is a representation compatible with human problem solving. The work domain is represented in a way that is similar to the mental model of the human operators. When the architecture is based on such a representation it is expected that the UAV will behave according to human expectations and become compatible with human interaction. The immediate benefits are that the abstraction hierarchy:

- provides a psychologically valid representation for goal directed problem solving.
- forms a common language for agents in a socio-technical domain.
- provides an informational basis for coping with unanticipated events.
- supports computationally economic problem solving.

Situation Awareness for UAVs

The next important question is: "what is situation awareness in a machine?" It is an interesting question because there is not a clear answer to what situation awareness is in a person. Before successfully integrating manned and unmanned flight it is necessary to have some understanding of how a machine can be aware of its situation and what that means. This paragraph is the result of a first assessment and explains what is though to be a useful path that will lead to UAV situation awareness.

The notion of situation awareness is hard to grasp, it is not tangible and at times seems to describe itself. As pointed out by Flach, Mulder and van Paassen (2004) it is important that we don't slip into using the description of the phenomenon as an explanation of the phenomenon.

To come more to grips with the concept of machine SA a comparison is drawn with the concept of safety. Safety is an important property of many systems we build, especially aircraft. Aircraft that are unsafe are not allowed to fly. It is a well defined property of the aircraft (by regulations) but nowhere can a component, a subsystem, a process or any 'box' be found in an aircraft that is labeled 'safety'. This is because safety is an aggregation of the properties of the components and their interactions. Safety has an abstract meaning and is not directly observable. When designing situation awareness the designer should not aim for building a box or a process that can be labeled 'situation awareness'. A UAV's situation awareness is, like safety, an aggregation of system properties, processes and their relation to the actual situation. It is reflected by the system's interactions with the environment, thus how it deals with the situation. The first step the designer should focus on is building an architecture for the system that allows it to understand the situation. Our first step is the domain representation as proposed in this paper.

Flach et al. (2004) state that an understanding of what is meant by the term 'situation' is essential for any progress toward a coherent theory of SA. The abstraction hierarchy is considered as a description of how experts organize or chunk complex information. In the same sense designing an understanding of the situation in the work domain is needed for any progress towards designing SA. And the abstraction hierarchy is proposed as a domain representation for understanding the situation; a means for the designer to chunk complex information in a way that is compatible with human reasoning.

Shared Situation Awareness

The term 'shared situation awareness' is used here to describe the capability of UAV, pilots and other operators to share their situation awareness. The importance of shared situation awareness to automation is discussed in relation to collision avoidance. Collision avoidance is very important for UAV operations because collision avoidance (sense and avoid) capability needs to be demonstrated before UAVs are allowed to fly in civil airspace. That the matter is more complicated than equipping UAVs with a Traffic Collision and Avoidance System (TCAS) is illustrated by what is referred to as the Ueberlingen midair collision. Nunes and Laursen (2004) describe the events of that night and identify a number of contributing factors, ranging from system malfunctions to human factors issues that took the safety redundancy out of the system. Under such circumstances it can be anticipated that some errors remain uncaught but what is striking is that TCAS, a system designed as a last safety measure to resolve a traffic conflict when all else failed, was unable to prevent a fatal accident. On board commercial jets TCAS interrogates the transponders of nearby aircraft. When a possible collision is detected one pilot is told to climb and the other to descend and thereby resolve the conflict. However, according to Nunes and Laurson (2004) TCAS itself was a contributing factor that led to the accident.

The Ueberlingen Accident

On the night of the 1st of July 2002 a midair collision took place above Lake Constance, Germany. The collision involved a Boeing 757 en route from Bergamo to Brussels and a Tupolev-154 that was flying form Munich to Barcelona. Both aircraft were equipped with the Traffic Collision Avoidance System (TCAS). The aircraft flew at the same altitude (FL 360) and their trajectories intersected at an angle of 90 degree above Lake Constance, they were on a collision course. Just seconds *before* TCAS gave both pilots a resolution advisory the air traffic controller at the Zurich Area Control Center contacted the T-154 and instructed the pilot to *descend* to FL 350 to avoid collision. Seconds later, TCAS detected the possible collision and instructed the Tupolev pilot to *climb* and the Boeing pilot to descend. The Russian Tupolev pilot received conflicting commands and decided to obey the air traffic controller and to ignore TCAS. The Tupoloev descended to FL 350 where it collided with the Boeing that had followed the TCAS advisory and also descended to FL 350. All 71 people were killed. TCAS conflict resolution is based on the assumption that both involved aircraft actually follow the resolution advisory. Free interpretation of the TCAS is incompatible with TCAS philosophy because it does not account for situations where one aircraft does not follow instructions as was the case in the Ueberlingen accident.

When there is a conflict between ATC and TCAS, European pilots are advised to follow the TCAS advisory. In contrast Russian pilots are trained to take both the ATC commands and TCAS advisory into account before making a decision. The British pilot of the B757 followed TCAS and descended to FL 350, and the Russian pilot of the T154 chose to ignore TCAS and follow the ATC command to descend to FL 350 as well. Why the Russian pilot took this decision at that time will remain unknown but it does point out that there must be arguments for pilots to assume that ATC is in control and has priority over TCAS. The fact that the Russian pilot had not contacted the air traffic controller about the conflicting commands suggests that these arguments might be quite strong. If it is indeed the case that pilots can have good reasons to believe that they should not obey TCAS the assumption that all aircraft follow the traffic resolution becomes unreliable. Unreliable because the parties involved based their situation awareness on different assumptions.

The air traffic controller did not know that the given command to descend was in conflict with the resolution advisory that TCAS issued seconds later. The Russian pilot in the T-154 probably thought that the air traffic controller was resolving the conflict and decided to obey the controllers command without confirming this. The British B757 did what made the most sense to him to avoid a possible collision and followed the TCAS advisory. The assumptions they made, made sense to their own understanding of the situation but were incompatible with one another.

The described TCAS problems can be translated into a lack of *shared situation awareness* as a contributing factor. What the TCAS contribution to the accident points out is that the situation awareness of one airspace user is not enough. The situation must be shared by all involved parties, they must have the same understanding of the situation and work domain; they must share situation awareness.

With respect to TCAS, improvements could be made to make sure that the controller has the same information as the pilots when a TCAS alert is triggered. One way of doing this could be to automatically inform ATC that a conflict is detected and that what advisories have been issued.

California Crisis

The above story cannot really be told without telling about how TCAS saved the day in a potential disaster unfolding in the southwestern U.S. skies on Tuesday 14th of September, 2004. The crisis occurred at the Los Angeles Air Route traffic Control Center in Palmdale California at around 5 pm. The center that is responsible for aircraft flying above 13000 feet suddenly lost contact with all 400 aircraft in 460 000 square kilometers of airspace over California and parts of Arizona, Nevada and Utah including the busy McCarren International airport in Las Vegas (Geppert, 2004). The cause was a software bug and left aircraft in the area without ATC guidance to keep them separated. Quick thinking controllers used mobile phones to alert other traffic control centers and the airlines that their aircraft were on a collision course but the real life saver was TCAS. Commercial jet pilots were able to avoid collisions by following the issued TCAS advisories. That evening no collisions took place despite the large number of aircraft involved.

This incident shows us that communication does not by definition enhance shared situation awareness. In this event the lack of communication gave the pilots no other choice but to rely on the TCAS resolution advisories for collision avoidance. Given the situation it was safe for pilots to assume that the other involved pilots relied on TCAS for collision avoidance as well and that it was their highest priority. The lack of communication made all pilots assume the same thing about their situation which resulted in a high degree of shared situation awareness and the safety of 400 airplanes.

Discussion and Future Work

The problem of UAV integration is a much larger problem than just fitting UAVs with clever 'sense and avoid' equipment. Because UAVs will be required to operate like manned aircraft, human factors is a core element of the integration process. UAVs need to have situation awareness like human pilots and they need to be able to share their world understanding with people. The abstraction hierarchy has been identified as a valuable framework for representing the work domain and the situation, i.e. the constraints shaping behavior. It is hypothesized that the abstraction hierarchy as a domain representation will form the basis for goal directed problem solving and dealing with unanticipated events.

Future research will focus on how the abstraction hierarchy can be formalized into software and used to reason about the world and engage in goal directed problem solving activities. The representation will be compatible with the human way of reasoning about the work domain. It can form the common language between multiple operators in the domain, including human (actors) and artificial operators (agents). When actors and agents make their decisions based on the same goal directed representation of the work domain they will be able to understand each other's behavior despite their different internal workings. Eventually this should lead to shared situation awareness which is a state in which multiple operators (artificial and human) have a great deal of similarity between their understandings of the situation.

Acknowledgement

The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024.

References

Airspace Integration Plan. (2004). *Airspace Integration Plan for Unmanned Aviation.* Office of the Secretary of Defense.

Flach, J., Mulder, M., van Paassen, M. M. (2004). The Concept of Situation in Psychology. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (p.42-60). Ashgate Publishing, Oxon (UK).

Geppert, L. (2004, November). Lost radio contact leaves pilots on their own. *IEEE Spectrum*, pp. 8–9.

Nunes, A. & Laursen, T. (2004) *Identifying the factors that led to the Ueberlingen mid-air collision: implications for overall system safety. Proceedings of the 48th Annual Chapter Meeting of the Human Factors and Ergonomics Society*, New Orleans, LA, USA.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. (1994*). Cognitive systems engineering*. John Wiley & Sons Inc.

Reynish, W. (2004, October). UAV's entering the NAS. *Avionics Magazine*, pp. 43–48.

Vicente, K. J. (1999). *Cognitive work analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEE Transactions on Systems, Man and Cybernetics*, 22 (4), 589–606.IEEE spectrum November 2004

UAV Task-Force. (2004). *UAV Task-Force Final report; A concept for European Regulations for Unmanned Aerial Vehicles (UAVs).* Joint Aviation Authorities and Eurocontrol.

FIGHTER PILOT TRAINEE RETENTION OF KNOWLEDGE AND SKILLS: AN EXPLORATORY STUDY

Dee H. Andrews
Patricia C. Fitzgerald
Air Force Research Laboratory
Mesa, AZ

Christopher B. Sinclair
Ian B. Farquharson
Advanced Training Systems International
Mesa, AZ

An exploratory study was conducted to investigate knowledge and skill retention of foreign military fighter pilot trainees with intermediate levels of flying experience. Twenty participants completed a standardized advanced skills fighter-training program that lasted about 10 months for the first class (n=12) and eight months for the second (n=8). Following flight training, the students engaged in non-flying duties (i.e., leave, English training classes). Members of the first class did not resume flying for a minimum of eight months; the second class returned to the simulator or the flight line within three months of completing initial training. Thus, two retention intervals were available for analysis. Analyses of instructor estimates of the students' skill and knowledge retention revealed significantly greater perceived decay among the students in the first class. Furthermore, the students in the second class were perceived to have been better prepared for their sorties than those in the first.

Introduction

Research psychologists have been examining the acquisition and retention of human learning for well over one hundred years. Learning acquisition has been extensively examined in many thousands of research papers. However, the retention of knowledge and skills acquired in the learning process has been less extensively studied and therefore less is known about the topic.

Pilots must learn a tremendous number of skills and considerable knowledge to be safe and effective. This learning takes place over many months or perhaps even years. While most pilot certification testing takes place soon after the initial learning occurs, the pilot may not be called upon to use many skills or pieces of knowledge until a considerable time after the initial learning takes place. The retention of skill and knowledge of pilots is the theme of the study reported in this paper.

Of the relatively few aviation learning retention studies that have been performed, most examined the retention of lower order skills such as procedures. As we explain in the literature review section of this paper, we have found few aviation learning retention studies that have examined higher order cognitive skills such as decision-making. This study examined retention of a variety of skills, both simple and complex, but we believe the most interesting findings relate to the complex cognitive skills necessary for basic fighter maneuvering and air combat.

One reason for this relative dearth of research into learning retention has to do with the difficulty of conducting such retention research, especially compared to what is required to investigate learning acquisition. Most human retention studies require the subjects to return to be retested days, weeks or months later. It is often difficult to entice all of the subjects to return for this retesting. Some reasons might be: subject leaves the local area, subject is too busy, the subject did not like the experiment in which they participated, or the subject simply forgets to come at the appointed time. Regardless of the reason, it can be difficult to get a complete sample of subjects to participate in the retention part of a learning study.

For this study the experimenters were able to avoid many of the problems usually associated with enticing retention subjects to return for the retention portion because the pilot subjects were enrolled in a military training program and they had to return as part of their military duties. In addition, the study had a unique advantage over other studies that have examined pilot learning retention because the pilots did not fly between their first training course and a seasoning course that was offered many weeks later. Typically, pilot trainees start flying operational missions shortly after their initial training is complete. Even if researchers wish to measure learning retention, the operational flying performed by recently graduated pilots serves to bias the retention measurements. That is, if the operational flying requires the pilot to use any of the skill or knowledge being measured in the retention study, the retention measures eventually taken are not true reflections of how much skill or knowledge decay that has occurred after the learning acquisition portion of the study.

A majority of the research concerning knowledge and skill retention has been conducted in the laboratory rather than in applied settings (Arthur et al., 1998; Hagman & Rose, 1983; Nembhard, 2000). Because the literature on natural tasks supports the contention that retention is stronger in this condition than for artificial tasks (Arthur, et. al., 1998), more research needs to be conducted in real world settings. This is important for the military because Reserve and National Guard units are often called to service with long periods of non-use of the skills required when deployed (Arthur et al., 1998). Furthermore, although retention research was conducted in aviation several decades ago, few recent research undertakings have addressed the issue. Finally, given the complexity of modern aviation systems, and the conflicting findings in the literature concerning the retention of complex tasks, it is necessary to readdress these issues.

Literature Review

The learning research literature records decades of studies examining the acquisition of knowledge and skills. However, by comparison to the acquisition literature, the literature on retention of skills and knowledge is relatively sparse (Hagman & Rose, 1983; Lance, Parisi, Bennett, Teachout, Harville, & Wells, 1998). Although the phenomenon has been studied for more than a century, the lack of regularities in the findings cause the construct to often be excluded from theories and models (Rubin & Wenzel, 1996). Despite the fact that retention has not been the subject of much research in aviation, empirical studies from a variety of domains have suggested a number of factors that have been associated with the decay of learned information and skills.

Retention Intervals

The retention interval is the period of time between the initial learning and the subsequent use of a skill or learned material. Research in which varying retention intervals were studied reported that retention decreased as the length of the interval increased (e. g., Adams and Hufford, 1962; Arthur, Bennett, Stanush, and McNelly, 1998). Fleischman and Parker (as cited in Prophet, 1976) found that participants trained on a flight simulator retained virtually all of their perceptual-motor skills after retention intervals of up to 24 months, after which decay was marked. Studies conducted by Bahrick (1984) and Bahrick and Phelps (1987) indicated that learned information started to decay shortly after it was acquired, but reached a plateau after five or six years.

Retention of Procedural Skills

The retention of procedural skills has received a great deal of research attention. In their meta-analysis of the literature on retention, Arthur, et al., (1998) found that procedural skills (e. g., pre-flight checks) were more prone to decay than continuous skills (e. g., tracking, flight control). Adams and Hufford (1962) reported nearly complete loss of procedural skills (i.e., a bomb toss exercise) following a 10-month retention interval.

In addition to being prone to decay, highly proceduralized tasks may have negative implications when an anomalous situation occurs. In their study on memory and cockpit operations, Nowinski, Holbrook, & Dismukes (2003), stated that when a habitual procedural task is delayed the typical cue is no longer present and the task may be forgotten, especially if the person is busy or tired.

Retention of Intellectual Skills

Although there is much research on the acquisition of complex intellectual skills, there is little literature on the retention of those skills. In their analysis on the retention of complex skills required to perform military tasks, Lance, et al. (1998) found that more complex skills were more likely to be forgotten than less complex skills, especially over long retention intervals.

In a study on the learning and retention of a complex industrial skill, Nembhard (2000) found that experienced workers learned and forgot faster than their inexperienced counterparts. As task complexity increased, however, the rate of decay evidenced for the more experienced workers decreased. Nembhard attributed the more robust retention rate to the better developed schemas of the experienced workers. Similarly, Sauer, Hockey, and Wastell (2000) conducted an experiment in which participants were trained to perform complex spacecraft life support control functions. They found that participants retained the skills acquired following an 8-month layoff, regardless of whether they received procedure-based training or system-based training in which a higher-order understanding of the system was fostered.

Practice

Investigators have found that retention is facilitated by spacing the initial learning over time, rather than by massing practice in a shorter time frame (Baddeley, 1999; Hagman and Rose, 1983). In a review of retention studies, Hagman and Rose found that spacing learning trials was most effective before the participant became proficient at the task. In addition, providing a greater interval between learning sessions was not as effective as spacing trials. During the early phases of learning complex skills such as flying, regular well-spaced lessons promote the acquisition of the requisite skills. Although the number of trials of any given procedure or maneuver are limited during each session, further practice occurs in subsequent lessons as the required skills are integrated.

Practice may also take place apart from the actual training conditions. Mental practice is "the symbolic, covert, mental rehearsal of a task in the absence of actual, overt, physical rehearsal" (Driskell, Copper, and Moran, 1994, p. 481). In their meta-analysis, Driskell et al. found that, although practice in the actual training condition was found to be more effective, mental practice enhanced retention for physical and cognitive skills, with a greater positive effect for cognitive tasks. The meta-analysis also supported the idea that mental practice was less effective when employed by novices. Finally, brief periods of mental practice were optimal; the benefits of the practice decreased as the practice period increased.

Methods

Participants

Twenty participant pilot candidates completed a standardized advanced skills fighter-training program in the A-4 aircraft that lasted about 10 months for one class (n=12) and eight months for a second class (n=8). Upon completion of the initial training program, the students in the first class engaged in duties that were not related to aviation (i.e., leave, English training classes) for a period of eight months. They then returned to the training facility for seasoning training. Students in the second class also had a break between initial and seasoning training, however, the retention interval was limited to three months.

The seasoning training included a combination of activities that were designed to enhance the retention of the previously learned skills and knowledge. Once the seasoning portion of the curriculum was completed, the students were introduced to new skills and knowledge.

Sixteen instructor pilots (IP's) were employed by a private commercial flight training company to instruct the students. All had previous fighter instructor pilot experience. For any given sortie, students were paired with an instructor based upon scheduling constraints. Thus, the students trained with a variety of instructors during the course of the program.

Retention Measurement Instrument

A paper questionnaire instrument was developed to obtain the instructor pilots' subjective assessment of the level of knowledge and skill retention exhibited by each trainee (see Appendix A). In addition, the instructors were asked to estimate the extent to which the student was prepared for the seasoning sorties. That question was asked so that the experimenters could make an estimate of whether student preparation contributed to the IPs estimates of retention. Both assessments were measured on a scale from 0 to 100, representing the percentage of retention and preparation. IPs also indicated whether or not they had instructed the student on the skill set in initial training. The assumption was that IP familiarity with the trainee from previous flights would likely lead to a higher estimate of retention. Finally, IP's indicated the sortie identifier, the date of the flight, and the student's class number (i. e., 1 or 2).

Procedures

Upon the completion of each flight during seasoning training, the IPs completed the instrument to provide an assessment of the student's retention and preparation for that flight. Five functions were included in the seasoning training and were evaluated for the present study. For each function, a series of re-familiarization sorties was flown. Transition training consisted of a series of flights that addressed aircraft handling and basic and aerobatic flight maneuvers. A series of simulator flights were conducted to practice emergency procedures. Instrument flight procedures, including basic instrument, radio, and navigation procedures, were also practiced in a series of flights. Basic and tactical formation skills were addressed in two- and four-ship formation flights. Finally, a minimum of ten training flights dealt with basic fighter maneuvers, including offensive and defensive maneuvers.

Results

Knowledge and Skill Retention

For the retention measure, a total of 102 usable IP ratings (64 for class 1 and 38 for class 2) were obtained for the sorties identified as the first flights using the skills associated with the function since initial training. Incomplete or illegible rating sheets were excluded from the analyses. T-tests were conducted to assess the IP's perceptions of the level of learning retention in the interval between the basic and the seasoning courses. Analyses of IP estimates of the students' retention for all sorties for each class revealed a significant difference between the classes ($t_{(100)} = -2.523$, $p < .05$), with greater decay perceived among the students in the first class.

Also of interest was the retention evidenced based on the type of function (e. g., emergency procedures, basic fighter maneuvers, formation). Due to the small number of IP evaluations for some of the function types for each class, statistical analyses were not conducted. To determine if there were evident trends between the classes, however, the data were plotted on a bar chart. As Figure 1 illustrates, retention was perceived by the IPs to be poorer for the first class in all phases of training with the exception of Formation.

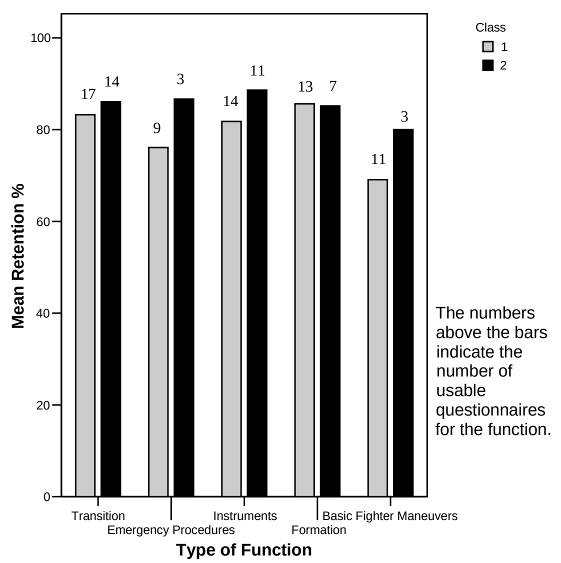


Figure 1. *Bar chart illustrating IP ratings of skill and knowledge retention by function.*

Student Preparation

Similarly, a difference was detected between the classes regarding the IP's assessment of student preparation for seasoning training ($t_{(82)} = -2.258$, $p < .05$). Students in the second class appeared to arrive better prepared than those in the first class.

The small number of assessments of student preparation for many of the sortie types also precluded statistical analyses at this level. The bar chart in Figure 2, however, illustrates a similar trend as was detected for knowledge and skill retention. Students in the second class were generally better prepared than those in the first for all function types except Formation.

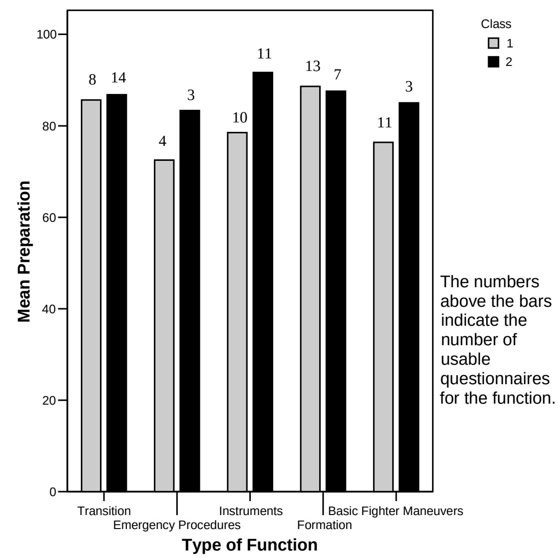


Figure 2. *Bar chart illustrating IP ratings of student preparation by function.*

IP/Student Training Continuity

T-tests were also conducted to assess differences in IP ratings based upon whether or not the pair flew together in initial training. Mean ratings of retention were not significantly different. Ratings of student preparation, however, were significantly different ($t_{(38)} = -2.653$, $p < .05$). IPs indicating that they flew with the student during initial training were more likely to rate preparation lower than those who had not flown with the student.

Discussion

It is difficult to design aviation learning retention studies that prevent the learning subjects from practicing their aviation skills between the initial learning events and the retention measurement. Pilots want to fly and look for every opportunity to do so. There is very little that will prevent them from flying, even if it is to advance the cause of science. This study took advantage of a mandatory aviation "grounding" of the learning subjects because they were not allowed to fly in the retention interval. For that reason the study is unique.

Due to the necessities of the aviation training program, the first group did not fly for eight months after their initial training course, and the second group did not fly for three months. Not surprisingly the IPs perception of the group with the shorter retention interval was that they retained significantly more skills and knowledge compared to the group with the longer retention interval. Clearly, the five extra months that the first group had to wait between their last flight in the initial training and the first flight in the seasoning training had a very deleterious effect on their overall performance.

An important question for future research is to examine whether the drop off in learning retention came fairly suddenly during the five additional months that the first group didn't fly, or whether the skill decay was consistently gradual across those five months. Co-authors of this paper, who are IPs instructing in the course described here and who have considerable Instructor Pilot experience, believe that the new learning decays at a fairly constant rate, and then suddenly drops fairly precipitously sometime between the three and eight month retention interval. Their experience, which is supported both by this study and by literature reviewed, is that procedural skills (e.g., emergency procedures) decay very rapidly, motor skills (e.g., landing skills) less rapidly, and higher order skills, such as decision making, decay with the greatest variability based on individual differences.

The students in the training program described here were not from the U.S., and English was a second language for them. The IPs in the program were convinced that language difficulty contributed to the skill and knowledge decay observed. It stands to reason that trainees who struggle with understanding concepts because their English language skills are deficient will suffer in both their acquisition of the skills and knowledge and perhaps in their retention of the skill and knowledge. The authors are not yet ready to ascribe retention difficulties solely to language problems. Since the IPs were only asked to rate the retention of the trainees in the two classes, and not to make judgments about the quality of their acquisition, it is difficult to know how much retention suffered compared to acquisition. The authors assumed that the trainees had reached at least the minimum criteria level in the acquisition phase of training since the trainees were all graduated to the seasoning phase. However, since actual acquisition levels were not measured as part of this study, it may be that language difficulties effected acquisition but not retention. The literature review did not reveal any studies that examined the impact of language skills on retention, but we suggest that this would be an interesting topic of research given the international nature of aviation training.

IPs in this study were asked to rate the flight preparation of trainees that the IPs flew with in seasoning flights. Not all IPs flew with all students in the acquisition stage either because of scheduling or because there were new IPs hired for the seasoning phase. The surveys revealed that IPs rated students with whom they had flown with in the acquisition phase of training as being less prepared for the seasoning flights than trainees they had not flown with in the acquisition phase. This finding seems counterintuitive because one might assume IPs would be somewhat biased toward students they had already instructed and would be more likely to give them higher preparation ratings. We believe that the counterintuitive finding might stem from a bias in the opposite direction from what we expected. That is, IPs had certain "pride of ownership" in the capabilities of the students they had previously trained and therefore had higher expectations for them in the seasoning phase of training. If that is true, we believe that the IPs were somewhat harsher in their judgment of the preparation of their former trainees than they were for students with whom they had not previously flown. Such a phenomenon would account for the low ratings for former students regarding their preparation.

Piloting skills and knowledge are prone to decay over time. We believe this study contributes to a fairly small body of literature that casts some light on this decay and retention phenomenon. The aviation research community can do the aviation industry a great service by continuing to conduct aviation skill and knowledge retention studies. Data gathered from these studies can be used to eventually build models of learning retention which would be of great value to those responsible for training and retraining pilots.

References

Adams, J. A. & Hufford, L. E. (1962). Contributions of a part-task trainer to the learning and retention of a time-shared flight maneuver. *Human Factors, 4,* 159-170.

Arthur, W., Bennett, W., Stanush, P., & McNelly, T. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11(1)*, 57-101.

Baddeley, A. D. (1999). *Essentials of Human Memory*. East Sussex, UK: Psychology Press Ltd.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, *113*, 1-29.

Bahrick, H. P. & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(2), 344-349.

Driskell, J. E., Copper, C., & Moran, A. (1994). Does mental practice enhance performance? *Journal of Applied Psychology, 79(4),* 481-492.

Hagman, J. D. & Rose, A. M. (1983). Retention of military tasks: A review. *Human Factors, 25(2)*, 199-213.

Lance, C. E., Parisi, A. G., Bennett, W. R., Teachout, M. S., Harville, D. L., Wells, M. L. (1998). Moderators of skill retention interval / performance decrement relationships in eight U. S. Air Force enlisted specialties. *Human Performance, 11*(1), 103-123.

Nembhard, D. A. (2000). The effects of task complexity and experience on learning and forgetting: A field study. *Human Factors*, *42*(2), 272-286.

Peck, A. C. & Detweiler, M. C, (2000). Training multistep procedural tasks. *Human Factors, 42*(3), 379-389.

Prophet, W. W. (1976). Long-term retention of flying skills: An annotated bibliography. HumRRO Final Technical Report FR-ED(P) 76-36 (ADA036114).Alexandria, VA: Human Resources Research Organization.

Rubin, D. C. & Wenzel, A. E. (1996). One-hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103*(4), 734-760.

Sauer, J., Hockey, G. R. J., & Wastell, D. G. (2000).Effects of training on short- and long-term skill retention in a complex multiple-task environment. *Ergonomics, 43*(12), 2043-2064.

Appendix A. Retention Measurement Instrument

This scale below presents a simple scale from 0 % to 100 %. For each flight we ask that you provide an overall assessment of how much of the skill set you believe the student has retained since the last time they used that skill set. That is, please give us an overall assessment of the amount of skill retention the student has maintained in the period between the last time they used the skill set and the flight you just finished with them.

Mission Number _________
Date __________________
Class 1 2

0 % = No retention at all of the skill set
100 % = Complete retention of the skill set

0% 10 20 30 40 50 60 70 80 90 100%

Did you instruct this student on this skill set in the initial training? Y N
How well prepared do you feel the student was for this sortie?

0 % = Not at all prepared for the sortie
100 % = Extremely well prepared for the sortie

0% 10 20 30 40 50 60 70 80 90 100%

Comments:

CAPTURING THE RESEARCH AND DEVELOPMENT PROCESS OF AVIATION SYSTEMS: CREATING A MULTI-MEDIA LIVING LEGACY

Anthony D. Andre
San Jose State University / NASA Ames Research Center
Moffett Field, CA 94035

Becky L. Hooey
San Jose State University / NASA Ames Research Center
Moffett Field, CA 94035

David C. Foyle
NASA Ames Research Center
Moffett Field, CA 94035

Too often, successful system development projects fail to leave a legacy of design transfer information, beyond providing access to the mere physical descriptions of the system, or the software code itself. Yet, information about high-level design decisions, assumptions, constraints, philosophies and methodologies is often sought after by system designers, engineers, and researchers alike. Such information is critical for facilitating an understanding of the design and evaluation decisions that underlie the final design. In contrast, published articles about a given complex system are usually limited to discussions of experimental results and in applicability beyond the academic and research community. This paper presents an argument for the development of an interactive multi-media design transfer library that provides a detailed legacy of the philosophy, design rationale and supporting data behind new aviation systems and conveys important guidelines, methodologies and "lessons learned" from the course of their research and development.

Introduction

To increase the efficiency and safety of surface operations, the Taxiway Navigation and Situation Awareness (T-NASA) cockpit display suite (see Figure 1), comprised of an electronic moving map (EMM) and a scene-linked head-up display (HUD was proposed, and then subjected to an extensive human-centered design and evaluation process over a 6-year period (Andre et al. 1998; Foyle et al. 1996; McCann et al. 1998; Hooey, Foyle and Andre, 2002).

During this period, nearly every type of research activity was performed, including:

- Jump seat field observations of pilots and air traffic controllers.
- Focus group studies with pilots and air traffic controllers.
- Studies using head and eye-tracking equipment.
- Low fidelity part-task desktop design concept studies.
- Medium-fidelity part-task simulation studies.
- Full-mission high-fidelity simulation studies.
- Flight tests in NASA's B757.

The focus of the studies varied as well, to include:

- Research to determine pilot information requirements during taxi.
- Research on user interface design options.
- Research to identify factors that contribute to current-day problems (safety/efficiency).
- Research comparing future operational concepts against current conditions.
- Research focused on crew roles and procedures.
- Research focused on systems integration issues.
- Research focused on near- vs. far-term technology assumptions.
- Research focused on benchmarking and quantifying safety and efficiency benefits of T-NASA.
- Research on usage characteristics.

Figure 1. *The T-NASA System.*

The Need for Design Knowledge Capture

Looking back on the T-NASA project, the research and development team realized that there was a vast quantity of information that could be passed on to manufacturers interested in the T-NASA system, regulatory agencies such as the FAA, aviation researchers and system developers, airlines and airline purchasing agents, and others outside of aviation who might generalize the philosophy, research approach and principle-based design techniques to their non-aviation product or system projects. Moreover, this information is not traditionally made available to those outside of the research and development team. For example, design concepts that were dismissed are rarely, if ever, discussed in publications or design specifications. Yet, that information, and specifically why a given design element was not deemed applicable or optimal for a given context, could be vital information to another researcher or developer, or to a regulatory agency.

Another common problem occurs when transferring software code. Often, those on the receiving end (manufacturers, system developers, etc.) forget that there is more to a system specification than just the software code behind the interface. Important design details, recommended procedures and other usage constraints are not contained within the code, and therefore can be easily ignored or misrepresented as the code travels through the development process.

Clearly, then, there is gap between what is typically published about the design or evaluation of a proposed system design and the information deemed necessary for facilitating an understanding of the critical design and evaluation decisions that underlie it. In an effort to both capture the activities and results of the T-NASA program and others like it, and to provide a useable form of traceability of the system philosophy, design guidelines, and research decisions, we argue the need for knowledge capture tools that can be used during the development process.

There are few tools in existence that purport to aid in the capture of design-relevant knowledge, and what tools do exist either focus purely on communications (e.g., the electronic cocktail napkin; Gross, 1996) or are used for the purpose of enabling people outside the project group to understand, supervise, and regulate what is done by the team (e.g., Gorry et al. 1991), or to secure intellectual property generated by the design team (Shipman & McCall, 1997). Further, they do not support real-time knowledge capture. Perhaps most telling is that few design teams make use of such tools.

While not the main focus of this paper, we advocate the future development of an easy-to-use, web-based, real-time knowledge capture or "design knowledge archive" tool; one that will capture, without undue effort on the part of the design team, high-level design decisions and rational associated with the design of complex aviation systems, as they are crafted. Such a tool would provide the underlying knowledge data base to support the automatic creation of an electronic, interactive multi-media design technology transfer library. The value and potential makeup of such a resource is described in the following section.

A Design Technology Transfer Library

The true amount of "data" and documentation that describes the research and development of a complex avionics system designed for human interaction can be daunting. In our initial concept for a prototype design technology transfer library, we have employed a familiar "ladder" metaphor. As shown in Figure 2 below, the user "climbs" the ladder, ending at the top shelf of the library with a description of the final design of the T-NASA system. The left side of the ladder presents the user with information specific to the development of the system, while the right side of the ladder presents the user with various categories of more generalized knowledge transfer information.

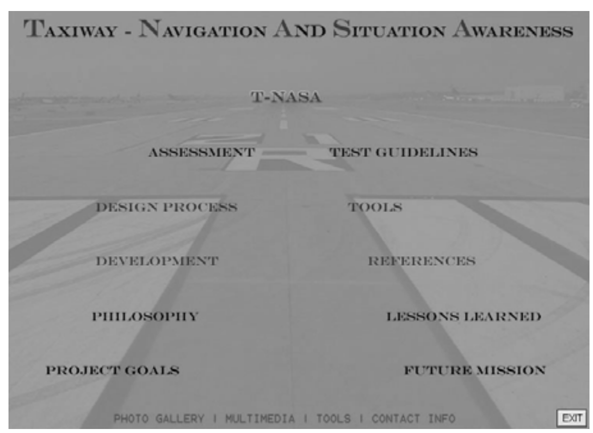


Figure 2. *Illustration of main menu category items from a prototype of the T-NASA design technology transfer library.*

The following is a brief description of the proposed purpose and content of each of these categories. The examples cited are specific to the T-NASA system and are intended only to illustrate the type of content that should be represented for any aviation system.

System Development Information

The categories of information related to system development are represented on the left side of the ladder in Figure 2.

Project Goals. To appreciate any system design one has to understand the project goals and objectives that the designers attempted to achieve. These goals and objectives may be defined by indices of safety, performance, capacity or usability, or specific use contexts, and may have derived from a government or industry program. For example, the main objective of the T-NASA system was to improve terminal area productivity in low-visibility conditions (Foyle et al., 1996). Design decisions were made based on this objective, which might have been different if, for example the goal was to improve safety in 'zero-zero' (no visibility) conditions. Specifically, for the former context we deemed augmented reality displays to be most appropriate, in which information is overlaid onto actual elements in the visual environment. In contrast, the latter context (no visibility) would require computer-generated virtual reality displays.

Clearly, then, without knowledge of the target goals and use contexts one could not understand, evaluate or appreciate the design of T-NASA. Worse still, the system could be adopted and used under circumstances for which it was never intended, creating safety hazards, or a failure to realize potential benefits.

Philosophy. Whether explicitly known to the designers or not, behind every design effort is an inherent design philosophy. This philosophy guides the design process and is the root of many design decisions. For example, a core philosophy of the T-NASA design was to support local control of the aircraft only with conformal, "head-up" information, while supporting global situation awareness with a head-down display (Foyle et al., 1996). Documenting, and communicating the design philosophy helps avoid "feature creep", and prevents future designers and developers from adding elements or modifying the design in a way that violates the original design philosophy.

Development History. Many end-users of this design transfer library may be interested in the development history of the system in question. Often, to better understand the ultimate design of a system, it is necessary to study the various incarnations it took during its development. This is a golden opportunity for the design team to explain and justify features and design elements that are NOT included in the final design. In fact, one could argue that it is often more informative to know why something was not included than to know why something was included.

For example, in the design of the T-NASA moving map, there was an active decision to NOT display taxiway centerlines in order to maximize eyes-out time and discourage the use of the map for local control purposes. Without documentation of this decision, and the rationale for it, future designers/developers could add a centerline without realizing the potential negative consequences.

In addition, systems engineers are often looking for information about a given system's hardware/software platform; information rarely specified in a human factors publication. Details regarding the assumptions that were made about data resolution, sensor reliability, and false alarm rates (as examples) are important to document. With rapid advancements in technology, it is very likely that what is considered a design constraint at the beginning of a design process is no longer a limitation by the time the system is fielded. This information would enable system engineers to differentiate between characteristics that were intended by design, or simply legacy due to (outdated) technology limitations.

Design Process. Capturing the design process and demonstrating a human-centered approach is recognized as an important element to document among the human factors community (e.g., Hooey, Foyle and Andre, 2001). Often, manufacturers or regulatory agencies are interested in the activities and process carried out to evaluate and/or validate the design. How were design requirements determined? How was the system tested? Were subject matter experts used to validate the proposed design? Was there a process to identify relevant procedural issues that might need to be addressed in order to accommodate the system? The processes that were engaged in to answer these questions can, and should be, articulated.

Evaluation/Assessment. Here, information on the assessment methods and data is found. Both quantitative and qualitative studies can be summarized, with samples of actual data, statistical analyses, etc. Documenting this information allows manufacturers, regulatory agencies, potential users, and purchasing agents to understand the extent to which the system has undergone a comprehensive evaluation process. For example, it is possible that a system demonstrates increased productivity, yet was never tested for safety impacts, or workload effects. Further, it is possible that a system was tested under nominal, or ideal operating conditions, yet was never tested under off-nominal or failure scenarios.

Without this form of documentation, it is difficult for various stake-holders to make informed decisions about adopting a system.

The System Design

In Figure 2 the final system design is represented by the T-NASA "shelf" at the top of the ladder. Here, the end-user would see the actual system design, be able to watch video of the system in action, and have access to an interactive design specification. The latter component could be presented in the form of an illustration with embedded hyperlinks that allows the user to hover over any design element and read a description and justification of that element.

In addition to design details, this category would also include information on usage assumptions, roles and responsibilities and assumed procedures. For example, information about usage assumptions can be helpful for future users of the system, those involved in developing training programs and standard operating procedures, and those responsible for integrating systems into future cockpits.

Knowledge Transfer

The categories of information related to knowledge transfer are represented on the right side of the ladder in Figure 2.

Test Guidelines. Beyond the data obtained from any given test or evaluation, it is often the case that useful methodological guidelines for testing similar systems or in similar contexts can be gleaned from the various research activities (Andre et al. 1998). As such, this section is devoted to conveying test guidelines, methods and best practices.

Tools and Techniques. Just as there are useful test guidelines to transfer, there are various tools and techniques employed by the design team over the course of the system's research and development that are useful to document. For example, a particular design technique (shadowing, perspective, transparency, etc.) or software program may have been used to render the specific look or behavior of a given interface element.

References. Most research and development efforts produce some amount of published material. Here, all references (and actual publication content) directly and indirectly related to the project are contained, ideally in an electronic form. Also this category could contain industry standards and guidelines that were used in the process.

Lessons Learned. All large-scale systems design projects are inherently educational in nature. Too often, the valuable lessons learned are not captured and transferred to future designers or engineers. This section provides an opportunity for the design team to communicate valuable information in perhaps a more personable form. Information on how system designers can best communicate design information to developers, or how to avoid feature creep are examples of useful lessons learned.

Future Mission. This section provides an opportunity for the design team to "close the loop" by indicating where the end-user might expect to see a commercial production of the system and/or future activities planned by the design team. In addition, insights into how the product may be adapted or useful for other contexts can be communicated.

Making it Interactive

Having the right information is one thing, making it easy, engaging and worthwhile to interact with is another. We advocate that the information contained in the library be presented in an interactive, multi-media format, making use of the latest software and audio-visual technologies, including images, sounds, animation and video.

Summary

Too often, successful system development projects fail to leave a legacy of design transfer information, beyond providing access to the mere physical descriptions of the system, or the software code itself. Thus, a gap exists between what is published or can be gleaned from looking at the final system design and the comprehensive library of knowledge, activities, guidelines and data often left to the memories of the design team. We argue the need for easy-to-use, real-time distributed software tools for capturing the knowledge and process behind the research and development of complex avionics systems. We advocate that the output of this tool be used as the input to an interactive, multi-media design technology transfer library, with the end-purpose of creating a detailed legacy of the philosophy, design rationale, development history and supporting data behind new aviation systems and conveying important guidelines, methodologies and "lessons learned" from the course of their research and development.

Acknowledgements

Funding for this project was supplied by the Airspace Operations HMP/HADMT Program, UPN 21-711-60-01.

References

Andre, A.D., Hooey, B. L., Foyle, D. C., and McCann, R. S., (1998). Field evaluation of T-NASA: Taxiway Navigation and Situation Awareness System. IEEE/AIAA Digital Avionics Systems Conference. Seattle, WA.

Foyle, D. C., Andre, A. D., McCann, R. S., Wenzel, E., Begault, D., & Battiste, V. (1996). Taxiway Navigation and Situation Awareness (T-NASA) system: Problem, design philosophy, and description of an integrated display suite for low-visibility airport surface operations. SAE Transactions: Journal of Aerospace, 105, 1411-1418.

Gorry, G. A., Long, K. B., Burger, A. M., Jung, C. P., & Meyer, B. D. (1991). The virtual notebook system: An architecture for collaborative work. Journal of Organizational Computing, 1(3), 233 – 250.

Gross, M. D. (1996). The electronic cocktail napkin – Computer support for working with diagrams. Design Studies, 17 (1).

Hooey, B. L., Foyle, D. C., & Andre, A. D. (2001). The design of aircraft cockpit displays for low-visibility taxi operations. In A. G. Gale (Ed.) Vision in Vehicles IX. Holland: Elsevier Science Publishers.

Hooey, B. L., Foyle, D. C., & Andre, A. D. (2002). A human-centered methodology for the design, evaluation, and integration of cockpit displays. In proceedings of the NATO RTO SCI and SET Symposium on Enhanced and Synthetic Vision Systems. September, 10-12, 2002. Ottawa, Canada.

McCann, R.S., Hooey, B.L., Parke, B., Foyle, D.C., Andre, A.D. & Kanki, B. (1998). An evaluation of the Taxiway Navigation and Situation Awareness (T-NASA) system in high-fidelity simulation. SAE Transactions: Journal of Aerospace, 107, 1612-1625.

Shipman, F. M., & McCall, R. J. (1997). Integrating different perspectives on design rationale: Supporting the emergence of design rationale from design communication. Artificial Intelligence in Engineering Design, Analysis, and Manufacturing (AIEDAM), 11 (2), 141-154.

MITIGATING WEATHER EFFECTS ON THE FUTURE NATIONAL AIRSPACE SYSTEM: THE INTEGRATION OF HUMAN FACTORS, DECISION SUPPORT AND DISPLAY TECHNOLOGIES

Anthony D. Andre, Ph.D.
Interface Analysis Associates
1131 S. De Anza Blvd.
San Jose, CA 95129
andre@interface-analysis.com
408 966 9355

Phil Smith, Ph.D.
Cognitive Systems Engineering Lab, Ohio State University
Columbus, OH 43210
smith.131@osu.edu

Amy Spencer, Ph.D.
Cognitive Systems Engineering, Inc.
Ostrander, OH 43061
amyspencer@columbus.rr.com

Jimmy Krozel, Ph.D.
Metron Aviation, Inc.
Herndon, VA 20170
krozel@metronaviation.com
503-274-8316

Weather is a major limiting factor in the National Airspace System (NAS) today, accounting for roughly 65% of all traffic delays. Because we cannot control weather and safety must be maintained in the presence of weather- related hazards, our ability to mitigate the effects of weather through advances in weather prediction, human factors, decision support tools, automation and display technology are critical to supporting the projected growth in air travel demand. This paper presents the core ideas, human factors approach, and initial display concepts for supporting all-weather operations in the future NAS, developed as part of NASA's Virtual Airspace Modeling and Simulation (VAMS) program.

Introduction

Weather is a major limiting factor in the NAS today, accounting for roughly 65% of all traffic delays. Because we cannot control the weather and because safety must be maintained in the presence of weather related hazards, our ability to predict the weather and how it influences air traffic are critical elements in designing the future NAS. As part of NASA's Virtual Airspace Modeling and Simulation (VAMS) project we have been developing concepts for mitigating weather effects, and thus restoring or increasing the NAS capacity, for the years 2020 and beyond.

The capacity of the NAS is ultimately limited by its ability to accommodate safe and efficient travel under *all* weather conditions. The key to greater capacity in the NAS lies in our ability to accurately predict and adjust the future state of the NAS on a timescale consistent with critical NAS response times. From a Human Factors perspective we have developed a triad of core ideas to represent our concept for increasing the NAS capacity in the context of weather. The core ideas are: 1) flexible traffic management, 2) shared situation awareness, and 3) coupled weather and traffic prediction. The "Core Idea Triad" is based on the philosophy that the optimal plans, strategies and responses for mitigating weather effects cannot be fully achieved without common situation awareness among different NAS users, coordination of traffic plans, and sufficient information sharing and transfer.

We have developed a set of scenarios that depict both current day and future concept operations in the context of capacity-limiting weather events, across different levels of scope (e.g., local weather events, ground vs. upper air weather, propagating weather events) and involving different sets of NAS users (pilots, ATC, traffic managers, dispatchers, etc.). Each scenario details the weather phenomena in question, how the weather impacts current-day operations, future roles and responsibilities, decision support tools (DSTs) and other user interfaces derived from our core ideas and concepts. Further, for each scenario we have developed a preliminary set of functional illustrations, which serve to demonstrate the information that a given user, or set of users, might have access to in the future NAS.

A Human Factors Approach

A major element of this project was the identification of key human performance objectives, listed below. Our belief is that these issues underlie many of the previous failed attempts to introduce automation and decision aiding to the NAS at a large scale.

- Improve the Distribution of Data and Provide Tools to Assist with its Use.
- Make Technologies Useful in Spite of their Brittleness
- Constrain the Solutions Suggested by a DST Based on Human Factors Considerations.
- Support Collaboration and Coordination among distributed NAS operators.

A Human-Centered Design Process

As part of the concept design process, we first developed a high-level human-centered design approach. This approach is represented by the following main human factors themes.

- Implement new technology to enhance performance while employing human-centered design techniques to support human decision making, keep operators in the loop and in control.
- Utilize communication and display technologies to share relevant information and perspectives between pilots, dispatchers, controllers and traffic managers.
- Help formalize and automate useful procedures carried out today in a manual, effortful and ad-hoc manner.
- Maintain current roles and responsibilities as much as possible, but support proactive problem solving through advanced technology, human-centered DSTs and shared awareness interfaces.
- Develop realistic solutions that can implemented in the near-term or in phases over time.
- Design distributed work systems and procedures in order to avoid excessive cognitive complexity and workload for any one individual.

Our User Interface Approach

Our approach to user interface design, which we intend to apply to all operators within the NAS, is to impact the user's ability to access, understand, integrate, and act on the variety of information sources, and to do so in support of both individual and group work, in a timely fashion and with undue levels of workload and stress. New and emerging sensor, algorithm and display technologies will be considered in our effort. Finally, our interface design approach is supported by the following design principles.

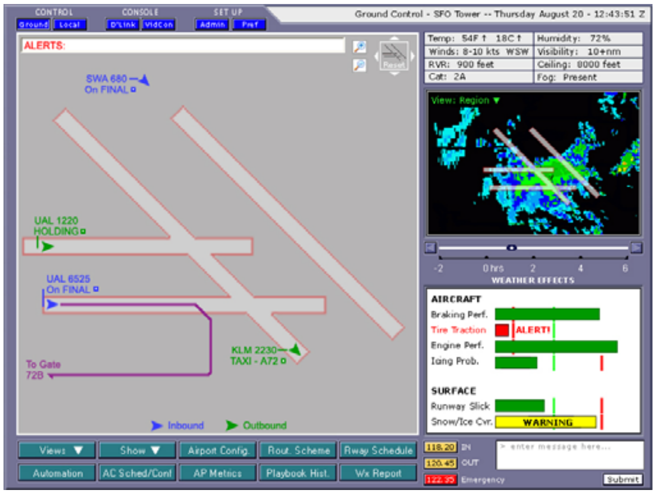
- Shared awareness – push relevant and context-sensitive, though not identical representations of, information to various NAS users towards facilitating collaborative decision making.
- User control/authority—support user, don't make decision for them.
- Transparency – allow the user access to the logic behind any calculation, algorithm or decision support solutions.
- Multi-modal – provide users with multiple information views or perspectives, taking advantage of different input and processing modalities.
- Collaborative – provide interfaces that make collaboration between NAS users efficient, easy and beneficial.
- Flexible – prevent automation and technology brittleness by allow the user to choose the parameters, to alter the logic, to add constraints not considered by the DST, to override automatically created values, and to adjust levels of uncertainty.
- Present Wx implications, not just data – provide the user with the implications or the effects of Wx, not just the data. In doing so, the interface is performing a common cognitive task for the human, that is, determining how Wx conditions will affect aircraft performance, airport surfaces, and other safety variables.
- Saliency – provide salient, at-a-glance indicators of overload, capacity loss, uncertainty, predicted effectiveness.
- Modeling and comparisons- provide the user with tools to model and compare various DST solutions, before selecting a specific initiative.
- User defined constraints – allow users to define and input constraints that may not be known to the computer system.
- Input of user priorities – allow timely and easy input and adjustment of user priorities.
- Visual modeling– provide layered visual representations of solutions, effects and Wx so that the user can easily see how the proposed

initiative will mitigate Wx and other constraints.

- History-provide the user with access to historical data (e.g., delayed or pop-up flights), success rates and system-derived estimates of the applicability of a DST solution to a given context or situation.

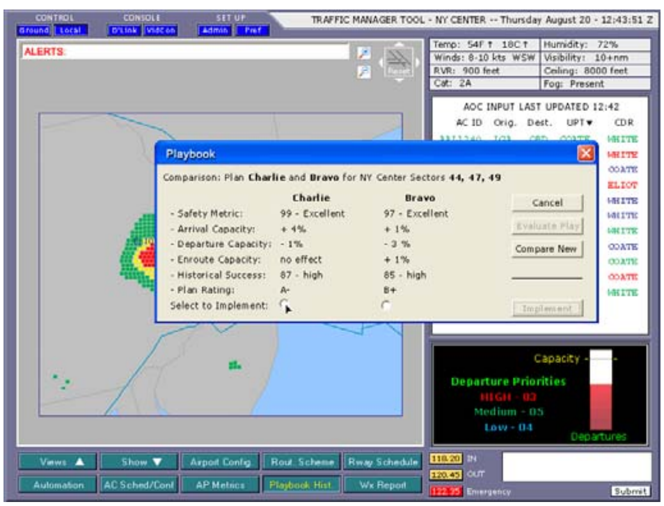
Concept Interfaces

On the following page we present some of the concept interfaces developed as part of this effort. They are already making a large impact on the aviation community and future plan for mitigating weather effects on our national airspace system.

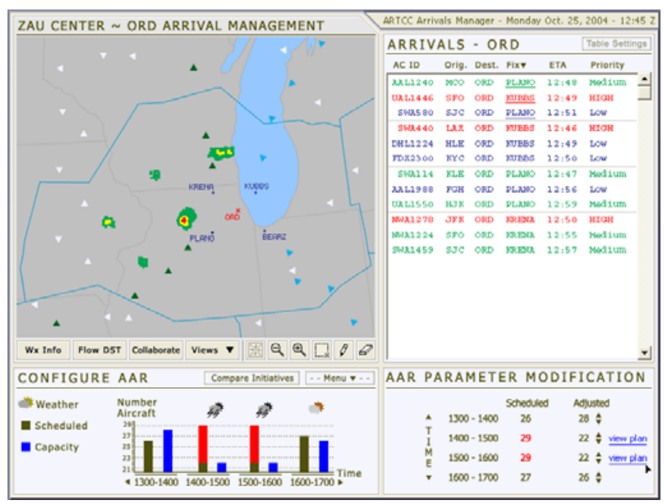


Future Ground Control SA Interface Showing Wx Effects

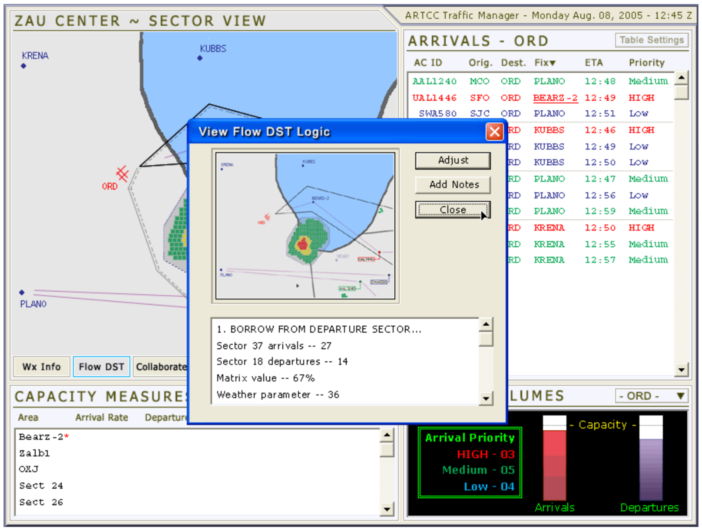
Future Ground Control SA Interface Showing Pilot View



Future Traffic Manager Interface Showing Plan Comparison



Future Center Controller Interface Showing Arrival Info



Future Center Controller Interface Showing DST Logic

Future Pilot Interface Showing Wx Optimized Route

References

Billings, C.E., "Toward a Human-Centered Aircraft Automation Philosophy", *International Journal of Aviation Psychology*, Vol. 1, No 4, pp. 261-270, 1991.

Hansman, J., and Davison, H.J., "The Effect of Shared Information on Pilot/Controller and Controller/Controller Interactions," *3rd USA/Europe Air Traffic Management R&D Seminar*, Napoli, June, 2000.

Krozel, J., Lee, C., and Mitchell, J.S.B., "Estimating Time of Arrival in Heavy Weather Conditions," *AIAA Guidance, Nav., and Control Conf.*, Portland, OR, Aug., 1999.

Penny, S., Krozel, J., Andre, T., *et al*, *Technologies Enabling All-Weather Maximum Capacity by 2020: Scenarios for the Capacity Increasing Concept*, Phase 2 Final Report, Technical Report, Metron Aviation, Inc., Herndon, VA, January 15, 2004.

LINE PILOT PERFORMANCE OF MEMORY ITEMS

Howard Au
Boeing Commercial Airplane Group*
P.O. Box 3707 MC 67-TC
Seattle, WA 98124, USA

* This research was paid for through out-of-pocket expenses by the author. It was conducted while the author was a student at Embry-Riddle Aeronautical University in partial fulfillment for the degree of M.S. Aeronautics.

An evaluation of Boeing 737 line pilot performance of memory items in 5 abnormal checklists was performed in a single-blind experiment using tabletop exercises at the crew base of a major U.S. airline. A study of 16 pilots shows that performance of memory items results in errors in identifying the failure, selecting the proper checklist to be completed, and checklist step errors.

Introduction

Some system failures that can occur on commercial airliners require flight crews to perform checklist steps from memory prior to referring to the checklist. These steps, called memory items or recall steps, are for time critical actions crucial to the safe continuation of the flight (e.g., preventing severe aircraft damage or crew incapacitation). Typically, line pilots do not study memory items except in preparation for a proficiency check (PC), usually every 6 or 12 months. They arrive for their evaluation prepared to be tested on the recall of the memory items. Their performance in these evaluations may not reflect their performance on the line, months after a PC.

This study examines whether line pilots are familiar enough with the memory items to perform all of them reliably, without prior knowledge that they will be evaluated. It was predicted that the performance of the memory items would show errors of commission, omission, and order due to the pilots' infrequent review of the memory items. This impromptu method of evaluation more closely resembles an unanticipated inflight emergency. This paper reviews some of the literature on performance under stress and then discusses the results pertaining to errors in identification of failures and errors in checklist selection. Although checklist step omission and order errors were observed, this paper will focus on the commission errors in the completion of checklist steps.

Human Performance Under Stress

An inflight emergency requiring timely action imposes a great deal of stress on the flight crew. Previous studies have shown that recall under high-stress conditions is more prone to errors than recall under low-stress conditions [8]. These errors, as they relate to checklist use, may include errors in identifying the abnormal condition, selecting the correct checklist, and errors of commission (adding steps or performing steps incorrectly), omission (missing steps), or order (completing steps in the wrong sequence).

Baddeley [1] presented a review of studies that included performance of deep-sea divers, combat aviators in actual combat, soldiers in simulated emergencies, and skydivers. These studies evaluated the performance of manual dexterity tasks, tracking tasks, and attention to peripheral cues. They showed that danger manifests itself in human performance through a narrowing of attention or through an increase in time to complete a manual dexterity task. The narrowing of attention can potentially lead to increased performance only if the task being performed is understood to be important. *However*, performance on tasks made to seem peripheral during an emergency can deteriorate [3]. Similarly, if the task is so complex as to require attention to numerous cues, the narrowing of attention will result in an inability to integrate relevant task information and an inability to conduct a proper assessment of the situation [6].

It is possible that training can mitigate some of these effects. *However*, even though pilots receive regular training in emergency procedures in simulators, that does not mean they are unaffected by the stress of an actual emergency. An emergency in a simulator is not perceived as life-threatening. If the pilot fails, the simulator can be reset for another attempt. Unless a pilot has had repeated experience in dealing with a truly dangerous emergency, performance in a real emergency could be similar to a novice. It has been shown that subjects are able to inhibit fear and prevent it from affecting their performance only if they are repeatedly exposed to a dangerous situation [1]. Due to the reliability of today's airliners, it is unlikely for the average airline pilot to have this kind of exposure in an airplane.

Stress Effects on Problem Analysis

It is possible that performance on infrequent tasks, such as identifying the root cause of multiple failures or shutting down an engine inflight, is affected more by stress than are common tasks. This is "an effect that has profound implications for the design of procedures to be used under the stressful conditions of emergency" [9].

This effect can sometimes be observed when people continue with a planned series of actions they are familiar with even when the actions appear unsuccessful or inappropriate. By acting before analyzing the situation, the operator may exacerbate the situation, which may induce more stress, and make it increasingly difficult to identify the original cause of the failure. This is related to an effect referred to as *confirmation bias*, where a person attends to cues that support a belief, and discounts cues that contradict the belief. Confirmation bias has been demonstrated in the use of automation and even in the diagnosis of everyday situations [4, 5, 7]. Other studies have shown that under stress, subjects are less effective and more disorganized at considering alternative solutions and incorporate less data in decision-making [6].

Stress Effects on Completion of Checklist Steps

Discussions with pilot participants in this study suggest that the requirement to perform certain actions from memory implies a sense of urgency in the performance of those actions. This introduces another potential source of error due to the loss of accuracy as speed is increased, an effect that is best described by the speed-accuracy operating characteristic (SAOC). The SAOC is a function that represents the inverse relationship between accuracy and speed. As the performance of a task requires more speed, accuracy is reduced until it approaches chance. If accuracy is excessively emphasized, then the time required to complete a task increases greatly with little improvement in accuracy.

Wickens & Hollands [9] summarize studies that demonstrate the effects of stress, induced by speed or by threat of bodily harm, on performance accuracy. For example, bomb-disposal experts performing under stress made more errors while working faster, and subjects who were threatened with the potential for electric shock gave up on problem-solving activities early.

Using an emergency descent as an example, an earlier study [2] showed that crews performing an emergency descent from memory took longer to descend than crews using the checklist. The difference in descent time resulted from omission errors by crews performing memory items. They occasionally omitted deploying the speedbrake, causing the airplane to descend slower. On the other hand, crews that performed the procedure by reference to the checklist did not make these errors, but took longer to complete the checklist. Regardless of the time required to read through the checklist, the crews performing the procedure by reference descended to a safe altitude in less time because of the use of the speedbrake.

The perceived requirement to perform checklist steps quickly from memory during high-stress situations is at odds with the need to perform those checklist steps accurately. There is a potential for loss of accuracy as the performance speed increases. Attempting 100% accuracy would require so much time to complete a checklist that other flying tasks would be disrupted. There is a tradeoff between getting the procedure done quickly, and getting it done while minimizing the possibility of error.

The following methodology seeks to identify examples of these errors in the flight operations domain. Even though inducing a level of stress similar to that of a real emergency was not possible in this study, it was hypothesized that errors of commission, omission, and order would still be observed.

Methodology

Participants

Sixteen 737 line pilots at a crew base of a major U.S. airline volunteered for the study. These pilots were already at the crew base either in preparation for a flight or returning from one. Participants were accepted without regard to experience level and participated in the study individually and not as a member of a two person crew. Pilots reported being trained in both the 737 Classic and 737 NG.

Procedure

In order to avoid any priming effects in the recall of their emergency procedures, subjects were not informed of the purpose of the research. They were instead briefed that:

- the research was on the suitability of the 737 alerting system,
- they would be asked to talk through five procedures, and
- the results from this study may be relevant to the design of a new alerting system in future airplanes.

A brief survey of experience was collected. This included data on total number of hours flown, their time in airplane type, flying time since last PC, and their crew position.

Subjects were seated in front of a poster of the flight deck. For consistency, a color poster of the 737 Classic flight deck was used. Five non-alerted abnormal procedures that contain memory items were used. They included aborted engine start, engine limit/surge/stall, rapid depressurization, runaway stabilizer trim, and dual engine failure.

The experimenter began each scenario by describing a normal flight situation, and then interjecting cues that suggest a particular failure. Subjects were asked to react to the cues as they would inflight, performing any procedures they felt were necessary. When responses to the scenarios seemed vague, the researcher probed the participants to encourage them to elaborate. The participants were provided with their airline Quick Reference Handbook (QRH), and were allowed to select the checklist they felt was most appropriate for the situation. Each session lasted approximately 30 minutes.

Results

Demographics

The participants in this study were 16 current line pilots at a major U.S. airline. Of those pilots, one was eliminated from the final analysis because he determined during the interview that an evaluation of the performance of memory items was the goal of the research.

	Total Time	Months Since PC	Time in Type	Weeks Since QRH Used
Mean	13,404	6	6,614	13
Standard Deviation	6,829	4	6,535	17
Minimum	4,500	0.5	400	1
Maximum	25,000	11	20,000	52

Table 1. *Demographics*

Data from the experience survey is presented in Table 1. Nine First Officers and six Captains participated. Two pilots incorrectly reported their total time and time in type, and their numbers were excluded. Seven pilots had prior military experience ranging from land and carrier-based fighters to large transports. Pilots who did not have military experience came from various corporate jets, commuter planes, other large commercial airlines, and corporate turboprops.

Checklist Selection Errors

When pilots were given an engine start condition with no oil pressure indications, four pilots initially chose the Engine Low Oil Pressure checklist. Upon reading that checklist, two of those pilots realized it was not appropriate for the situation, and correctly selected the Aborted Engine Start checklist. One pilot reported that there was no checklist needed, and that a maintenance call would be the only action required after completing the engine shutdown. The remaining 10 pilots correctly referenced the Aborted Engine Start checklist (Table 2).

Checklists selected	# of pilots
Aborted Engine Start	10
Engine Low Oil Pressure	2
Engine Low Oil Pressure > Aborted Engine Start	2
None	1

Table 2. *Aborted Engine Start Checklist Selection*

The Engine Limit/Surge/Stall scenario had the lowest identification rate (Table 3). Only two pilots referenced the correct checklist. One of those two selected the Engine Fire/Severe Damage/Separation checklist first. The remaining pilots referenced various checklists, including Engine Fire/Severe Damage/Separation, Engine Failure/Shutdown, and Engine Overheat.

Checklists selected	# of pilots
Engine Limit/Surge/Stall (Correct)	1
Engine Fire > Engine Limit / Surge / Stall (Experimenter prompted the correct checklist by saying the engine was "surging")	1
Engine Failure	6
Engine Fire	4
Engine Overheat > Engine Fire	1
Engine Overheat	1
Engine Overheat > Engine Failure	1

Table 3. *Engine Limit / Surge / Stall Checklist Selection*

The remaining three scenarios had few checklist selection errors. One pilot selected the Auto Fail/Unscheduled Pressurization Change checklist during a rapid depressurization. Another pilot performed the Stabilizer Out Of Trim checklist in the runaway stabilizer scenario.

Checklist Step Errors

The majority of checklist step errors occurred during the completion of the dual engine failure memory items. Many of those were commission errors. These included:

- bringing the thrust levers back to idle before attempting to restart the engine,
- advancing the thrust levers as the engines failed in an attempt to get them to restart,
- starting the APU to try an assisted start,
- waiting three seconds to attempt a restart after shutting off the fuel,
- placing the ignition selector to both, and
- using engine anti-ice (Figure 1).

→ Ignition Selector................................Both
→ Thrust Levers..............................Advance
→ Engine Start Levers............................Idle
Engine Start Switches.......................Flt
→ Turn around
→ Thrust Levers..................................Close
→ Engine Anti-ice.....................................On
Engine Start Levers.....................Cutoff
EGT decreasing:
→ Wait three seconds:
Engine Start Levers..........................Idle
→ APU..Start
If EGT exceeds 950°C:
→ Allow engines to overheat
Repeat above steps
→ Attempt restart one at a time

Figure 1. *Dual Engine Failure Commission Errors. Bold items indicate the correct steps. Arrows indicate all additional steps performed by the 15 pilots.*

In the rapid depressurization scenario, two pilots included additional steps:

- verifying the engine bleeds were on, and
- closing the bleed air isolation valve (Figure 2).

→ Engine bleed switches..........................On
→ Isolation valve..................................Close
Oxygen masks & regulators....On/100%
Crew communications............Establish
Pressurization mode selector.........Man
Outflow valve................................Close

Figure 2. *Rapid Depressurization Commission Errors.*

Four pilots made commission errors in the completion of the runaway stabilizer trim checklist by attempting to activate the electric trim switches in the direction opposite the runaway. One of those four pilots stated that he would also attempt to engage a different autopilot in the hopes that it would not experience the same malfunction (Figure 3).

Control column.....................Hold firmly
Autopilot (if engaged)...........Disengage
→ Electric trim in opposite direction
→ Engage other autopilot
If runaway stabilizer continues:
Trim cutout switches..............Cutout
Trim wheel....................Grasp & hold

Figure 3. *Runaway Stabilizer Commission Errors.*

Discussion

Checklist Selection Errors

When presented with cues to an abnormal situation, pilots sometimes omit a thorough analysis of the situation. This became evident through previous observations of pilots performing abnormal procedures in simulators and anecdotal evidence. The pilots in this study demonstrated a tendency to fixate on the most prominent cue and perform the checklist appropriate to that cue. However, a thorough analysis of the situation can reveal that the single most prominent cue does not always lead the pilot to the correct checklist.

There were 23 checklist selection errors. With the following three exceptions, the errors appear to be caused by the pilots' fixation on a single cue. Experimenter error in describing the rapid depressurization failure to one pilot gave the impression that the cabin altitude began to stabilize at approximately 12,000 feet, which led him to the Auto Fail/Unscheduled Pressurization Change checklist. Another error was due to a pilot's belief that no checklist was required for an aborted engine start. Finally, one pilot referred to the Dual Engine Failure checklist as the Engine Inflight Start checklist, but performed the correct memory items.

The remaining 20 checklist selection errors appear to be caused by pilots fixating on a single cue, and performing the checklist that appears most related to that cue. For example, in the aborted engine start, the cues given to the pilots were the continued illumination of the LOW OIL PRESSURE light and no oil pressure indication. Four pilots stated that, given those cues, they would complete the Low Oil Pressure checklist.

Two of those pilots realized the Low Oil Pressure checklist was inappropriate by considering the reasonableness of the checklist steps they were reading. The checklist directed the pilots to the Engine Failure/Shutdown checklist, which is meant for an inflight engine shutdown. A shutdown of an engine on the ground is simpler than a shutdown inflight and these pilots determined that irrelevant steps such as: starting the APU, maintaining fuel balance, and preparing for a single-engine landing, indicated they were in the wrong checklist. However, one pilot who entered the Engine Failure checklist from the Low Oil Pressure checklist did not consider the appropriateness of the checklist steps he was reading, and showed a tendency for perseveration. He went so far as to complete the Engine Failure checklist, reading aloud and bypassing irrelevant steps to complete the only step required to actually shutdown the engine while on the ground.

In the engine limit scenario, the 14 subjects who did not select the correct checklist instead performed the checklist that most closely reflected the cue they said was the most important. One pilot initially selected the Engine Fire/Severe Damage/Separation checklist, but turned to the Engine Limit/Surge/Stall checklist only after the experimenter said the engine was "surging". The term "surging" was not used as a cue in any other scenarios. Pilots who were primarily concerned by the abnormal "popping" or "banging" noises referenced the Engine Fire/Severe Damage/Separation checklist, stating that they believed the noises suggested severe engine damage. Pilots who considered excessive exhaust gas temperature (EGT) to be more important completed checklists related to overheat conditions. The pilot who referenced the Stabilizer Out Of Trim checklist in the runaway stabilizer scenario did so because he believed the STAB OUT OF TRIM light would be illuminated.

Checklist Step Errors

There appear to be consistent patterns in the observed checklist step errors. Many of the commission errors appear to result from the pilots' creativity in dealing with an abnormal situation. It was observed that many pilots perform steps in addition to what was required based on their understanding of how the airplane systems functioned, even though their understanding of the systems may be incorrect. Some pilots explained that the performance of some additional steps occurs because of knowledge of the intricacies of a complex system gained over years of experience or knowledge of common and simple failure modes, which are not addressed in the checklist. This may resolve the situation without the need for a checklist. In other cases, an incorrect or incomplete understanding of the system may lead pilots to perform additional steps that delay the completion of steps necessary to resolve the situation, or that may exacerbate the condition.

The pilots' creativity in dealing with certain situations was most evident in the dual engine failure scenario, which had the highest number of commission errors. A possible explanation was apparent in the pilots' response to this scenario: a desire to "do whatever it takes" to resolve a serious situation. Their perception was that this failure was so severe that they would exercise their authority as pilots, beyond what is written in the checklist, in an attempt to get an engine running, regardless of the consequences. Some pilots' willingness to allow the engines to exceed EGT and overheat, contrary to the guidance in the checklist, demonstrated this belief.

Most errors of commission were intended to troubleshoot the failures, such as: advance the thrust levers, verify the start levers are at idle, turn around to exit the heavy rain that caused the failure, and manually select both igniters. This last step demonstrates a misunderstanding of the ignition system. By correctly completing the recall item in the checklist, both igniters were automatically energized.

When the situation called for a shutdown of both engines, two pilots performed the additional step of delaying 3 seconds between restart attempts. They explained that this stemmed from a folk belief carried over from their military background that additional time was needed for excess fuel to clear the engine before attempting a restart.

This disposition towards creative troubleshooting was also seen in the Runaway Stabilizer Trim and Rapid Depressurization checklists. Errors of commission included moving the electric trim switches in the opposite direction and engaging the other autopilot. One pilot reported that he had experienced a runaway stabilizer in the past, and activating the electric trim switches stopped the runaway. This is an example of a pilot's knowledge of the failure modes of a complex system that could resolve the situation without using a checklist.

The rapid depressurization scenario showed that some commission errors, such as closing the isolation valve and ensuring the engine bleeds are on, would not exacerbate the situation, but would not be beneficial either. They would simply delay the completion of the necessary steps. Moreover, the manual closing of the isolation valve demonstrates a lack of understanding of the bleed air system. This step is not required because the valve is already closed during its normal operation.

On the other hand, some commission errors aggravated the situation. An example was seen in some pilots' willingness to allow the engines to overheat while restarting after a dual engine failure. The consequence of the overheating could be engine damage and a true engine failure, instead of the original problem of a temporary flameout due to an environmental condition such as heavy rain, resulting in no engine damage.

Conclusion

The results demonstrate that pilots have difficulty identifying the cause of the failure and selecting the correct procedure. After identifying the situation, knowledge of the appropriate memory items is such that pilots commit errors in recall even during unstressed conditions with a poster of the flight deck for context.

None of the five failure scenarios in this study had a distinct indicator light that would annunciate the condition. Pilots were forced to analyze the cues and determine the appropriate procedure. This is an uncommon and involved task, and not performing it may force pilots to complete only those tasks they are familiar with, such as following an illuminated LOW OIL PRESSURE light to the Low Oil Pressure checklist during an aborted engine start, or fixating on abnormal engine noises and performing the Engine Fire/Severe Damage/Separation checklist, instead of the more appropriate Engine Limit/Surge/Stall checklist.

The observed checklist step errors showed that pilots commit a number of errors. The majority of the commission errors were steps performed by pilots to resolve a failure based on their knowledge of the airplane systems. Some of these commission errors demonstrated a misunderstanding of how the systems in the 737 functioned. Other errors were a result of either knowledge gained during a real experience in the past, or a belief carried over from previous organizations and airplanes, which may no longer be applicable.

Implications

Even though the method used in this study did not induce stress, it allowed for an evaluation of the pilots' knowledge of the memory items without prior preparation. Pilots generally perform well during their PCs, and possibly better than inflight, because they expect an evaluation and can prepare for it. Pilot performance observed in this study may be closer to that in an inflight emergency, in which the pilots are unprepared to perform their memory items.

Clearly, an inflight emergency places a pilot under a great deal of stress. Based on the literature review, it can be inferred that errors similar to those observed here may occur inflight during an actual emergency, and may even occur more frequently due to increased stress. Conducting a similar study in a full-flight simulator may provide a level of stress similar to what is experienced in a real emergency. The results obtained from a simulator could be a more realistic representation of the results obtained inflight.

Acknowledgments

The author wishes to thank Barbara Holder and Randy Mumaw for their valuable guidance throughout the course of this project. The author also wishes to thank reviewers of drafts and the pilots who assisted by reviewing and commenting on research methodology, which greatly influenced the quality of data collected. The author also wishes to thank those who gave their time to allow us to gain some insight into the challenges they face in the cockpit.

References

1. Baddeley, A. D. (2000). Selective attention and performance in dangerous environments. *Human Performance in Extreme Environments*, 5(1), 86-91. (Reprinted from Baddeley, A. D. 1972. *The British Journal of Psychology*, 63(4), 537-546.)
2. Hamman, W. R. (n.d.). Quick Reference Checklist (QRC). *United Airlines validation study*. Colorado: Author.

3. Idzikowski, C., & Baddeley, A. D. (1983). Fear and dangerous environments. In D. H. Holding (Series Ed.) & G. R. J. Hockey (Vol. Ed.), *Stress and fatigue in human performance* (pp. 123-144). New York: John Wiley & Sons.

4. Mosier, K. L., Skitka, L. J., Dunbar, M., McDonnell, L. (2001). Aircrews and automation bias: The advantages of teamwork? *International Journal of Aviation Psychology*, 11(1), 1-14.

5. Reason, James. (1990). *Human Error*. New York: Cambridge University Press.

6. Salas, E., Driskell, J. E., & Hughes, S. (1996). Introduction: The study of stress and human performance. In E. A. Fleishman (Series Ed.) & J. E. Driskell & E. Salas (Vol. Eds.), *Stress and human performance* (pp. 1-45). New Jersey: Lawrence Erlbaum.

7. Skitka, L. J., Mosier, K. L., Burdick, M., Rosenblatt, B. (2000). Automation bias and errors: Are crews better than individuals? *International Journal of Aviation Psychology*, 10(1), 85-97.

8. Thompson, L. A., Williams, K. L., L'Esperance, P. R., & Cornelius, J. (2001). Context-dependent memory under stressful conditions: The case of skydiving. *Human Factors*, 43(4), 611-619.

9. Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance*. New Jersey: Prentice Hall

HUMAN CENTERED DECISION SUPPORT TOOLS FOR ARRIVAL MERGING AND SPACING

Vernol Battiste, **Everett Palmer**, **Walter Johnson**, **Nancy Smith**
NASA Ames Research Center

Tom Prevot, **Joey Mercer**, **Stacie Granada**, **Nancy Johnson**, **Quang Dao**, **Paul Lee**
San Jose State University
NASA Ames Research Center

A simulation of terminal area merging and spacing with air traffic controllers and commercial flight crews was conducted. The goal of the study was to assess the feasibility and benefits of ground and flight-deck based tools to support arrival merging and spacing operations. During the simulation, flight crews arrived over the northwest and southwest arrival meter fixes and were cleared for the flight management system arrivals to runways 18 and 13 right. The controller could then clear the aircraft to merge behind and space with an aircraft on a converging stream or to space behind an aircraft on the same stream of traffic. The controller remained responsible for aircraft separation. Empirical research was performed to assess air and ground tools and the effects of mixed equipage. During the all tools conditions, 75% of the arrivals were equipped for merging and spacing. All aircraft were ADS-B equipped and flew charted FMS routes which were coordinated based on wake turbulence separation at the arrival runway. The aircraft spacing data indicate that spacing and merging were improved with either air or ground based merging and spacing tools, but performance was best with airborne tools. Both controllers and pilots exhibited low to moderate workload and both reported benefits from the concept.

Introduction

At the core of the concept of Distributed Air-Ground Traffic Management is the idea that National Airspace System (NAS) participants can be information suppliers and team members who collaborate at all levels of traffic management decision making (Raytheon ATMSDI, 2003). One such concept and the focus of this paper is Concept Element 11 (Terminal Arrival Self-Spacing for Merging and in-Trail Separation).

Sorensen (2000) characterizes the CE 11 approach process as involving one of three operational modes. Each mode possesses potential benefits but also presents significant operational and technical challenges. These modes are: Free Maneuvering, Merging, and Spacing. During *Free Flight Maneuvering*, equipped aircraft can design their own direct path within a defined approach corridor (not under investigation in this study). *Merging* occurs when an equipped aircraft is delegated the responsibility for adjusting in-trail position behind the designated lead aircraft approaching from another stream; finally, the *Spacing* concept is one in which an *equipped aircraft is cleared to maintain a specified temporal position from a designated lead aircraft.*

The objective of CE 11 is to minimize the in-trail spacing buffers between terminal area arriving aircraft flying under instrument meteorological conditions (IMC). CE 11 utilizes time-based, in-trail spacing to take advantage of the natural spacing compression of arriving aircraft as they decelerate in preparation for landing (Abbott, 2002). To support the transition of responsibility for maintaining the desired spacing interval, from the controller to the flight crew, advanced ATM technologies (decision support tools – DST) were developed for both controller and flight crews (Granada, Dao, Wong, Johnson, Battiste, 2005).

In a previous study of merging and spacing, NASA ARC researchers employed a human-in-the-loop simulation with pilots and controllers, and tested time-based merging and spacing. Results of this study highlighted the need for clear delegation of responsibilities and unambiguous procedures under a variety of operational scenarios. Specifically, controllers were unclear about pilots' separation responsibilities. This ambiguity was particularly apparent when aircraft were spacing less than the assigned interval but still further than the legal separation requirement. Results of a follow-up study at NASA ARC reflected the progress made through the development of tools and procedures. When given the choice of issuing a spacing clearance to equipped aircraft, the TRACON controllers opted to provide the clearance about 85 percent of the time. This finding suggests that controllers were comfortable with the tools and procedures, and confident with the ability of pilots to accurately self-space (Lee, et al., 2003).

During an operational evaluation of in-flight spacing and merging, display integration was identified by flight crews as an issue when spacing information was presented on the NAV Display (ND). The FAA Safe Flight 21 operational evaluation data collected from flight crews identified display integration, clutter, and heads-down time as important display integration issues (Cieplak, Hahn, and Olmos, 1999).

The Flight Deck Display Research Group at NASA Ames has designed a suite of tools which should enable operators to safely and efficiently perform the necessary merging and spacing tasks essential to the success of the concept. In this report, we focus mainly on the evaluation of the flight deck DST. However, some discussion of the controller tools and tasks are necessary to set the context in which the flight deck tools were evaluated. The cockpit situation display (CSD), which is presented on the ND, includes a 3-D cockpit display of traffic information (CDTI), and the merging and spacing tools (FDDRL, 2004). The CSD integrates information derived from the spacing algorithms with traffic position, aircraft identification and intent to present a display of the current and predicted traffic situation (see Figure 1). Armed with this information and tools, flight crews were allowed to perform airborne merging and spacing operations when cleared to do so by the controller. This paper also examines the feasibility of the merging and self-spacing concepts from the flight deck perspective under mixed traffic conditions, where only some of the aircraft were equipped for self-spacing and merging. See Callantine, Lee, Mercer, Prevot and Palmer (ATM-2005) for CE-11 ground side results.

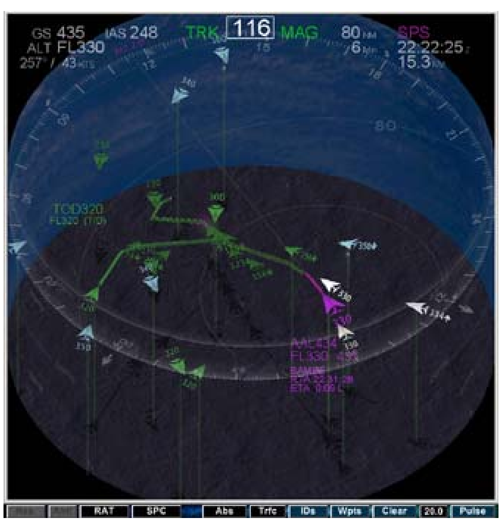


Figure 1: *3-D Cockpit Situation Display*

Methods

Pilot Participants

Nine air transport and/or commercial rated pilots and four certified professional controllers participated in the study. Pilots had an average of 10,405 flight hours and 3,912 hours in glass cockpits. All flight crew members were familiar with the advanced 3-D CDTI display system and received 2 days of training on the merging and spacing task and procedures. Four full performance level controllers with TRACON experience manned the feeder and final control positions in dual TRACON operations.

Experimental Conditions

Four experimental conditions were created to examine pilot and controller performance: No Tools, Ground Tools only, Air Tools only, and Air & Ground Tools. Data was collected from thirty two trials, with eight trials per condition. To assess the operational feasibility of the concept from the flight deck perspective, the following items were assessed: assigned vs. achieved inter-arrival spacing, usability/usefulness, flight crew workload, and safety. Additionally, pilots were asked to provide comments on the issue of call sign confusion when multiple aircraft IDs (call signs) are used in a single transmission. Post run and simulation questionnaires were used to assess concept feasibility and display usability.

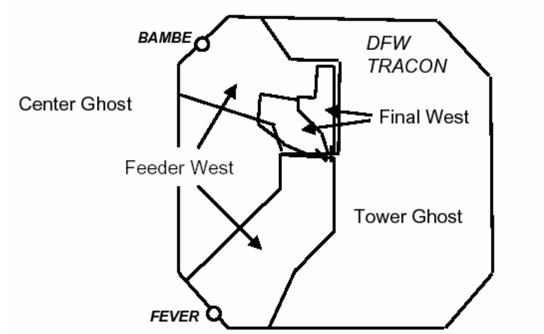


Figure 2: *DFW TRACON Airspace.*

Airspace and Controller Tasks

Controllers pairs (feeder and final) managed the western portion of the Dallas Fort Worth TRACON airspace. The feeder controller initially cleared the aircraft for either the Fever or Bambe FMS arrival, and if applicable, to follow a lead aircraft to 18R (see Figure 2). The Final controller managed the merge between the two arrival streams, which were procedurally separated by 1,000 feet at the GIBBI intersection.

Controller Display and Tasks

Controllers utilized a wake-vortex aware arrival schedule, which computed estimated times of arrival for runway 18R. In the conditions with ground tools, merging and spacing information was incorporated into each aircraft's data tag. For example, as illustrated in Figure 4,COA 538, a B733, landing 18R, assigned to follow BAW 601 80 seconds in trail and is currently 69 seconds in trail. Additionally, the spacing circle provides relative information about the spacing goal (see Figure 3).

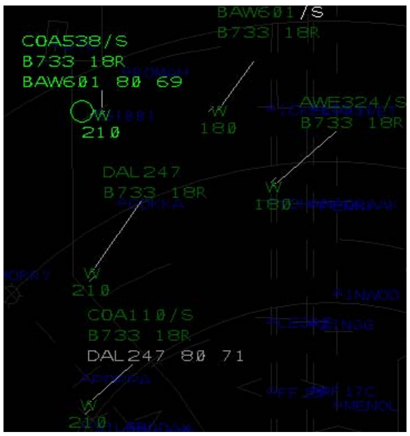


Figure 3: *Controller Display with merging and spacing tools.*

Roles and Responsibilities

Controllers were responsible for separation at all times. Flight crews could be cleared to merge behind then follow a lead aircraft on a conjoining route or to follow an aircraft on the same route. Controllers could cancel a spacing clearance at any time.

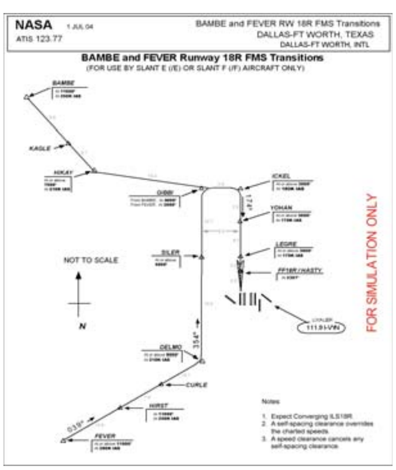


Figure 4: *FMS transitions to runway 18R - Streams merged at GIBBI.*

Procedures

Each aircraft started the scenarios 15 to 40 nm from the BAMBE or FEVER meter fixes. Upon entry, pilots were cleared to fly an FMS arrival route (see Figure 4) and were instructed to allow their aircraft to fly and descend along the FMS arrival path, even if Ownship seemed to follow another aircraft too closely – i.e., they did not adjust speed or altitude unless commanded by the air traffic controller (ATC). Pilots checked in with controllers when they received a data link clearance or at 5 nm from the meter fix. Pilots were instructed to expect spacing clearances any time after reaching the meter fix. Controllers issued clearances to merge and follow or follow behind a designated lead aircraft. Controllers utilized normal controller procedures – radar vectors, "direct to", speed and altitude – to manage the unequipped aircraft. The pilots utilized the airborne spacing tools and procedures to implement the assigned spacing command.

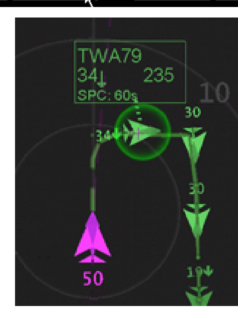
Pilot Clearance and Tasks

ATC provided clearances such as "Continental 538, merge behind then follow Speedbird 601– 80 seconds in trail," or "Continental 538, follow Speedbird 601 – 80 seconds in trail." Pilots read back the clearance and engaged self-spacing; see flight deck procedures below. If the algorithms did not command appropriate speeds based on the spacing setting, pilots were asked to disengage spacing and inform ATC that they were unable to space.

Tools for Merging & Spacing Operations

If a merging and spacing clearance was assigned, the flight crew followed the steps listed below using a mouse to position the curser:

1) Pilots first clicked on the Spacing button on the CSD tool strip.

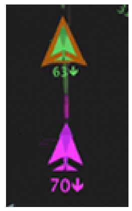


2) Pilots then selected the assigned lead aircraft by clicking on its symbol within the CSD. In this case, TWA79 was selected.

3) The spacing interval specified by ATC was then entered. To increase the spacing interval, pilots right-clicked on the seconds

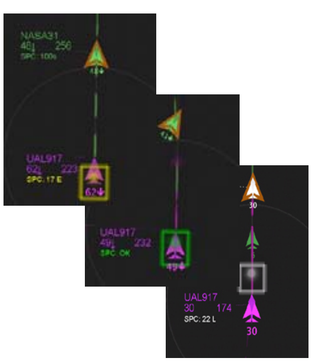
(**Sec:XX**) button; to decrease the interval pilots left-clicked the seconds button.

4) Pilots then clicked the start button on the CSD tool strip, which is located next to the "seconds" button in the figure above. Pilots were informed they would need to wait for the spacing algorithms to initialize. When the spacing algorithm was initialized (i.e., ready to engage spacing) the upper left corner of the CSD displayed a message indicating the spacing status. Also, the lead aircraft became highlighted in orange.



5) Finally, to engage the auto throttles, pilots selected the SPC button on the MCP. This activated the algorithm to begin commanding the proper speeds (via the auto throttles) to move the aircraft towards the spacing goal.

6) When the spacing is engaged and active, feedback is provided at the upper left corner of the CDTI.



Visual feedback regarding Ownship spacing status was provided via a color-coded "spacing box." The color and location of the spacing box reflected Ownship position relative to the assigned temporal spacing value. That is, if Ownship was given an assigned spacing value of 100 seconds and was more than 10 seconds ahead (e.g., the aircraft is currently at 83 seconds), the spacing box was depicted as yellow and Ownship appeared slightly ahead of the box. When Ownship was less than 10 seconds ahead or less than 20 seconds behind the assigned spacing value, the spacing box was depicted as green, and Ownship appeared inside the box. Finally, if Ownship was more than 20 seconds behind the assigned spacing value, the spacing box was depicted as white and Ownship appeared behind the box.

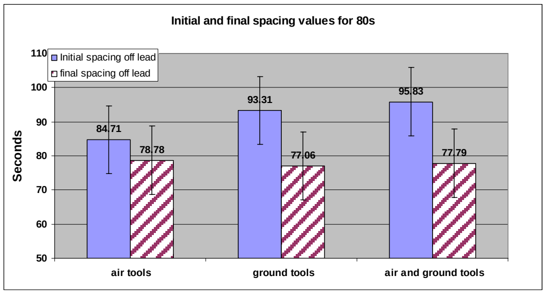
Simulation environment

The simulation study was conducted utilizing three fully integrated NASA ARC research laboratories/facilities: the Airspace Operations Laboratory (AOL), Flight Deck Display Research Laboratory (FDDRL), and Crew Vehicle Systems Research Facility (CVSRF). See DAG-TM, 2003 for a full description of each laboratory.

Results

This section presents the results of the Merging and Spacing operation at the 80 and 100 second intervals. Additionally, data on the efficiency of the merging and spacing operation, flight crew workload, safety and acceptability are described. Participating flight crews conducted 256 total approaches, 32 in each condition.

During the No Tools condition flight crews followed ATC guidance as they would today, thus no relative spacing and merging data are reported. Of the remaining 128 runs in the air tools and air and ground tools conditions controllers assigned spacing to the flight deck 116 times.



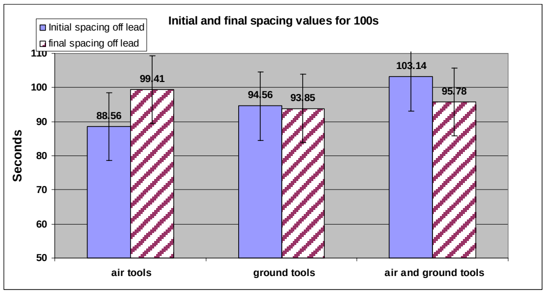


Figure 6 and 7: *Initial and final spacing intervals for 80 and 100 seconds (mean and standard deviation).*

Figures 6 and 7 show the spacing intervals data from the start of spacing and merging and/or spacing until spacing was discontinued at or near the final approach fix or by the controller. These graphs illustrate that, overall spacing performance was improved for All Tools condition and that performance was best in the Air Tools only condition (mean 78.8 and 99.4, respectively), followed by Air and Ground Tools (77.8 and 95.8), and finally Ground Tools (77.6 and 93.8). However, these trends were not significant (p >.05). Additionally, the expected improvement in spacing performance with air and ground tool was not found. However, controllers preferred to conduct spacing operations with only ground tools. They suggested that conducting merging and spacing operations when flight crews were managing spacing added additional variability and made it difficult to manage unequipped aircraft.

Spacing efficiency

From the flight deck perspective, a measure of efficiency was related to when aspacing and merging clearance was issued by the controller. If the clearance was issued early in the approach, the flight crews had more time to set up the systems and manage progress toward the spacing goal. If the clearance was issued late (i.e., near the base to final leg of flight), then this task may interfere with other tasks that require completion before landing. A t-test was conducted to examine this notion. The pilots' data was split into three groups; early, middle, or late approach clearances. A one-sample t-test was used to compare the three groups relative to the 80-second spacing goal. Results indicated that the early or mid approach groups did not significantly differ from the 80-second spacing goal ($p > .05$). However, when the spacing clearance was issued late, the spacing performance did significantly vary from the 80-second spacing goal, $t(22) = -3.33$, $p < .01$, indicating a decline in spacing performance (see Figure 8).

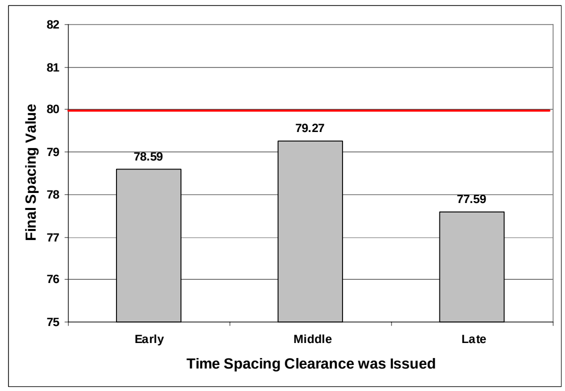


Figure 8: *Spacing performance with early, mid and late spacing clearance.*

	Air		Ground		Air/ Ground		None	
	M	SD	*M*	SD	*M*	SD	*M*	SD
Peak Workload	2.56	.69	2.25	.67	2.40	.72	2.23	.61
Overall Workload	2.34	.65	2.22	.67	2.28	.69	2.21	.63
ATC Communication	2.51	.75	2.52	.61	2.31	.58	2.53	.74

Table 1: *Crew workload and communication by conditions.*

Workload, Communication and Usability

After each approach, pilots entered a workload rating reflective of their perceived workload for the run using a modified NASA Task Load Index (TLX). There were a total of 32 trials in which the pilots provided workload data. The TLX rating scales were modified to include a peak workload assessment and an estimated communications workload relative to normal operations. Additionally, each rating was based on a Likert scale format that had "Normal Ops" as the median rating of 3 on a scale of 1 to 5, with a rating of 5 for “High” workload. This method was not used to suggest that “Normal Ops’ represents a medium level of workload, but it provided a familiar baseline for the participants. For this report only the peak, overall, and ATC communication workload values are presented.

The mean Peak Workload value was 2.45, SD = .72, the mean Overall Workload was 2.25, SD = .66, while the mean ATC Communication Workload was 2.39, SD = .57. Across all conditions flight crews’ ratings were relatively similar. The mean workload ratings were subsequently examined for each of the four conditions (Air tools, Ground tools, Air and Ground tools, and No Tools) separately. Table 1 includes the mean workload values for Peak Workload, Overall Workload, and ATC Communication Workload by each of the four conditions. As the table shows, flight crews rated the workload of the merging and spacing task below that of normal operations for all conditions (where normal operations was represented by a value of 3). The table also shows that crews rated communication workload lowest in the air/ground tools condition, suggesting that when both pilots and controllers have supporting tools, communication may be reduced.

ATC Clearances

An issue, which has stimulated considerable discussion over the past few years, has been the potential call sign confusion that may result in a DAG-TM environment. Specifically, the DAG-TM environment requires the use of two aircraft call signs in a single voice transmission. The concern has been that pilots may become confused by the use of two call signs and, at a minimum, may need to ask ATC to repeat the clearance. In a worst case scenario, the potential confusion could result in a pilot accepting a clearance that was meant for another aircraft. Of course, this worst-case scenario could lead to an accident or incident. An important finding in the present study was that, of 323 spacing and merging clearances, neither pilots nor controllers reported a single instance of “call sign confusion.” Flight crews reported that with the inclusion of flight ID and the pulse predictor (c.f., Granada et.al., 2005) on their CDTI, they were able to identify their prospective lead aircraft and to anticipate the ATC clearance.

Acceptable		
	Merging and spacing task	4.8
	Head-down time	4.0
	Display symbols	3.4
	Symbol Color	4.3
Useful		
	Information in aircraft data tag	4.0
	Accept spacing clearance based on CDTI data only	4.8
	Accept visual approach clearance based on CDTI only data	3.7
Safety		
	CDTI improves flight safety	4.3
	Enhances safety of merging and spacing	3.8

Table 2: *User Feedback on display, tools and conce*[cut off] *(N=10; 1 = not acceptable, useful and safe, 5 = very acceptable, useful and safe scale).*

As Table 2 shows, flight crews found the tools, display features and the concept acceptable, useful and safe. Also, these ratings suggest that the flight crews may be willing to take on additionally responsibility.

Discussion and Conclusions

Based on flight crew and controller performance, comments and also their interactions with the tools and procedures, the concept of merging and self-spacing during arrival and approach seems feasible. Pilots consistently rated the flight deck tools favorably in terms of usability, usefulness, and rated the CSD favorably in terms of situation awareness. Generally, pilot and controller workload ratings were moderately low during spacing and merging operations. Workload differences between tools conditions were relatively small for pilots, and when spacing clearances were issued early or at the mid point of the approach, pilots had little difficulty achieving the spacing goal. In this study, pilots and controllers generally disagreed as to the best time for the spacing clearances to be issued; however, the controllers were only beginning to develop strategies for how to best utilize this new tool. Finally, this study did identify a number of issues from the flight crews’ and controllers’ perspectives that need to be addressed in future research.

References

Abbott, T. (2002). *Speed Control Law for Precision Terminal Area In-Trail Self Spacing.* NASA Technical Memorandum 2002-211742. Hampton, VA: National Aeronautics and Space Administration.

Cieplak, J., Hahn, E., & Olmos, B. (2000). Safe Flight 21: The 1999 Operational Evaluation of ADS-B Applications. *Paper presented at the 3rd USA/Europe Air Traffic Management R&D Seminar*, Napoli, Italy.

FDDRL. (2004). 3D-CDTI User Manual. http://human-factors.arc.nasa.gov/ihh /DAG_WEB/ TRG/study_docs.htm.

Granada, S., Dao A.Q., Wong, D., Johnson, W.W., & Battiste, V. (2005) Development and Integration of a Human-Centered Volumetric Cockpit Situation Display for Distributed Air-Ground Operations. *Submission to the 2005 International Symposium on Aviation Psychology,* Oklahoma City, OK.

Lee, P., Mercer, J., Prevot, T., Smith, N., Battiste, V., Johnson, W., Mogford, R., & Palmer, E. (2003). **"**Free Maneuvering, Trajectory Negotiation, and Self-Spacing Concepts in Distributed Air-Ground Traffic Management." *In Proceedings of the 5th Eurocontrol / FAA ATM R&D Conference*. Budapest, Hungary.

Raytheon ATMSDI Team. (2003). *NASA Ames Research Center DAG-TM CE 6/11 Simulation, Final Report.* NASA Ames Research Center. Moffett Field, CA.

Sorensen, J. (2000). *Terminal Arrival: Self-Spacing for Merging and In-trail Separation.* Contractor Report NAS2-98005 RTO-41. NASA Ames Research Center. Moffett Field, CA.

EMERGING TECHNOLOGIES FOR DEPLOYABLE AIRCREW TRAINING

Dr. Benjamin Bell
Associate Chief Scientist
CHI Systems, Inc.
bbell@chisystems.com

Flight training devices commonly used for aircrew training offer high-fidelity simulation, wide field-of-view projection, detailed terrain, and realistic instrumentation and controls. Despite the significant investment needed to acquire and operate them, high-fidelity training devices enjoy widespread acceptance among end-users, air carriers, and military organizations.

Advances in computer simulation technology have helped reduce hardware requirements while providing software tools for scenario authoring, entity creation, performance assessment, and briefing/debriefing. A consequence of improved simulation tools is that training devices can be developed for a broader range of computational platforms, from very high-fidelity dedicated systems to desktop flight simulators running on standard PCs.

Choosing the appropriate technologies requires careful consideration of operational factors including training requirements, end-user priorities, logistics, cost, size/composition of the crew being trained, and the role of the instructor (if any). Current training systems research and development is evaluating the training value derived from current simulation technologies while exploring new approaches to extend the reach of simulation-based training.

Several promising research efforts are underway to develop training technologies that include intelligent tutoring, realistic synthetic entities, speech dialogue, performance assessment, and automated after action review. But a critical factor in the success of a training device remains the match between the fidelity of the simulation and the training requirements. For training airmanship and tactical air combat maneuvers, *physical* fidelity is a highly relevant property. Training that focuses on judgment and decision-making requires simulated environments that possess a high degree of *cognitive* fidelity. For training that emphasizes team skills, simulations should provide realistic *social* fidelity. If a focus of the training is radio communications, a simulator ought to provide a measure of *dialogue* fidelity.

This panel explores the range of issues surrounding how best to harness the power of emerging simulation technologies to create sophisticated aircrew training systems while at the same time carefully maintaining the consonance between the simulation and the training need. Each member of the panel possesses extensive experimental and applied backgrounds in modeling and simulation, training, or cognitive science, and has current responsibility for directing aviation training research and development. Each panelist will present a perspective on which approaches are likely to meet with success, and will share recent experiences from specific aircrew training initiatives. Following the presentations, a discussant will compare and critique the panelists' viewpoints and invite comments and questions from the audience.

Planned Panelists (one of these will be a discussant; which one is still TBD)

Dr. Benjamin Bell (organizer)
Associate Chief Scientist, CHI Systems, Inc.
bbell@chisystems.com

Dr. Winston Bennett
Lead, Training Systems Technology and Performance Assessment Team
Air Force Research Laboratory, Warfighter Training Research Division
winston.bennett@mesa.afmc.af.mil

Dr. Jared Freeman
Vice President for Research
Aptima, Inc.
freeman@aptima.com

Dr. Steven Hampton
Associate Provost for Graduate Programs and Research
Embry Riddle Aeronautical University
hamptons@erau.edu

Dr. Florian Jentsch
Director, Team Performance Laboratory
University of Central Florida
fjentsch@pegasus.cc.ucf.edu

Ms. Maureen Bergondy-Wilhelm
Head, Training Technology Development Branch (Air 4961)
Naval Air Systems Command, Training Systems Division
maureen.bergondy@navy.mil

COGNITIVE TUNNELING, AIRCRAFT-PILOT COUPLING DESIGN ISSUES AND SCENARIO INTERPRETATION UNDER STRESS IN RECENT AIRLINE ACCIDENTS

Meredith A. Bell[1]
Department of Psychology
University of Central Florida

Eugenio L. Facci
Department of Psychology
University of Central Florida

Razia V. Nayeem
Department of Psychology
University of Central Florida

By building upon a number of accident reports and on cognitive psychology literature, this paper addresses the effect of stress on the reasoning abilities and on the perceptual processes of pilots. We examine several cases, including American Airlines 587 (New York, 2001), United Airlines 173 (Portland, 1978), KLM 4508 (Tenerife, 1977), Northwest Airlines 6231(Thiells NY, 1974), and Eastern Airlines 401 (Everglades, 1972), in which pilots have, or may have, contributed to an accident by incorrectly interpreting the unfolding scenario, and specifically by disregarding alternative interpretations of the unfolding scenario. While current research efforts have yet to provide guidance on how to successfully handle the problems discussed in this paper, examination of prior accidents may shed some light on the issue.

Introduction

Operator performance under stress is a topic that has been under scrutiny for decades. In an environment in which operational settings contain a range of stressors, it is important to understand the effects of these stressors on operator performance in order to compensate for the possible decrements that result. One specific operational setting which has been prominent in this field is the aircraft cockpit. Aircraft pilots are faced with an array of stressors, ranging from environmental stressors to which they are routinely exposed and trained to endure, to those associated with emergency situations. Although there has been extensive research in the field, creating situations in which equivalent stress is produced has proven quite difficult if not impossible. The levels of stress induced, though probably lower than those with which a pilot would be faced during an emergency, have proven successful in detecting effects on pilot/operator performance. Some conflicting data have resulted; however, enough studies have come to the same conclusion that stress can have negative effects on operator/pilot performance in several different modes (Wickens, et al., 1993), (Driskell, et al., 1999), (Barnett & Wickens, 1986 as cited by Wickens et al., 1993). Research has indicated that the arousal of stress may severely disrupt a pilot's ability to objectively evaluate the situation with which he is faced. Specifically, cognitive tunneling can occur. Cognitive tunneling is a phenomenon in which a pilot will not adequately perceive all pertinent information because of filtering based on preexisting expectations, initial impressions or other undefined factors. This increases the likelihood that sensorial stimuli and alternative scenario interpretations would only be considered if consistent with these pre-existing expectations. While there is limited experimental data on the effects of stress equivalent to that experienced during a flight emergency, aviation safety records provide examples of this phenomenon. There are limitations to studying cognitive tunneling through post hoc analysis of accidents: it is subject to 20/20 hindsight and provides limited basis for generalization and prediction (Wickens et al., 1993). However analysis through experimental research has its disadvantages as well: it is difficult to achieve the level of stress that operators would face in an emergency. That is why it may be useful to explore this phenomenon through both approaches. In the following sections, the effects that stress has been found to have on operator/pilot performance will be examined, and several flights resulting in aircraft accidents will be reconstructed to explore the effect that stress had on the respective flight crews. The goal of this paper is to fill in some of the gaps left open by research with the archival analysis of previous accidents.

[1] All authors contributed equally to this paper. Order of authorship was determined randomly.

Cognitive Tunneling and its Cohorts

Cognitive tunneling has been recognized for years as a threat to operators who are faced with difficult decisions in the midst of an emergency. It is one of the many theories that surround decision making under stress. It does not act alone, however. Cognitive tunneling, sometimes referred to as attention narrowing, works in conjunction with several other phenomena that may collectively severely affect an operator's decision-making. The effects of these phenomena are cumulative, and as each occurs, the detrimental effect of the previous is often increased. Consequently, the operator is typically left with a decreasing amount of relevant information with which to work, more puzzling phenomena, and an increasing load on her/his cognitive processes. Wickens et al. (1993) present a model which provides an effective illustration of the stages of the decision-making process and the effects that stress has on each. This is the paradigm through which the phenomenon will be examined.

Cue Perception

The first stage, cue perception, is the first phase affected by stress. In most operational environments, there are numerous cues that must be considered when performing the required tasks. When operators are faced with a stressful situation, there is a tendency for the reduction in number of cues that are sampled and therefore perceived (Wickens & Flach, 1988). This selective allocation is referred to as selective attention, and while it is beneficial from a time/resource management point of view, operators sometimes allocate their attention poorly. There are many factors that can influence the distribution of attention, including reliability of the cue, saliency of cue, past experience with the cue, operator's pre-existing expectations and potential outcomes associated (Wickens & Hollands, 2000). Hence, pilots will pay for instance most attention to blinking lights or sounding alarms or to gauges confirming their initial interpretation of a problem. However, there are contradictory stances on the impact that stress has on selective attention. It has been theorized that stress actually improves selective attention. Chajut & Algom (2003), along with several others, have found that by imposing stress on an operator, she/he is better able to focus on the target task and rule out irrelevant cues. This is not entirely contradictory to the theory being presented herein. Stress decreases attention resources, and, therefore, greater efficiency is achieved by not sampling irrelevant cues and focusing on those deemed relevant to the problem. However, some of the cues deemed irrelevant are sometimes relevant, and the "efficiency" achieved comes at the price of embracing an incorrect interpretation of the unfolding scenario.

Working Memory

In the next phase of the model, the hypotheses stored in long-term memory are accessed and those assumed to be relevant are placed in working memory for evaluation. Additional narrowing can occur at this phase. While several hypotheses are stored in long term memory, only those associated with the preexisting expectations and the presumed problem will be retrieved, omitting several possible alternatives. Operators will then likely fixate on these hypotheses. Also, a function of working memory is the evaluation of action outcomes which are also retrieved from long-term memory. Increased stress places greater demands on this already "fragile" working memory, which degrades decision making (Wickens, et al., 1993). Hence, when pilots are faced with emergency situations, instead of evaluating all hypotheses learned in training and through experience, and thoroughly evaluating each, pilots evaluate the hypothesis they believe to be relevant with limited consideration of action outcomes.

The Cohorts

There are many issues that work hand in hand with cognitive tunneling to add to the effects of stress. One of these partners is confirmation bias, which occurs when an operator forms a premature hypothesis and seeks out cues and information to support solely this hypothesis (Wickens & Hollands, 2000). Many times operators believe they know what is causing the problem before they have even considered all the options, and instead of collecting information to test all of the hypotheses, they collect only information pertaining to the presumed cause. The operator is then left with a small set of information with which to work. When this information does not add up and confirmation of the hypothesis is not possible, further potential confirming cues are usually sought, while disconfirming information is usually not considered - the operator tends to perseverate. Belief perseverance, another collaborator, takes place when a person continues with a familiar plan of action even though it is fruitless (Ross & Lepper, 1980).

Expert vs. Novice

There has been extensive research in the area of cognitive tunneling regarding expert/novice differences. Deitch (2002) found that one of the most obvious differences in this area was cognitive mapping, where experts had more sophisticated cognitive maps and could relate their maps to more specific scenarios than novices. Other studies have found that there is a difference between experts and novices for instrument fixation, a task thought to be linked closely to cognitive processes (Harris, Tole, Stephens, & Ephrath, 1982). Additionally, in some tasks, experts even utilized different brain regions than novices (Peres et al., 2000). However, Guilkey (1997) determined that when pilots are faced with especially cognitively exhausting problems, flight time (experts vs. novices) is not a good predictor of performance. Results from this study indicated that no matter the strategy used, experts' performance was equal to novice performance. From these differing results, one can see that there are still many areas in cognition with respect to expert/novice differences to be researched; however, the evidence points to the problem of cognitive tunneling as one which faces both novice and expert pilots alike.

In the following section, the above theories are expounded upon through exploration of their presence in several aircraft accidents. By illustrating the existence of these phenomena in reality, not just in a simulator setting, compelling support is provided for these theories.

Aircraft Accidents

In this section, we examine five accidents in which the phenomenon of cognitive tunneling most likely played a significant role.

- American Airlines 587, Belle Harbor, NY, November 12th 2001 – As the flight was cleared for takeoff, the first officer - the flying pilot - asked the captain whether he thought sufficient distance had been allowed from the preceding plane, a large Japan Airlines aircraft, in order to avoid wake turbulence. The captain stated "aah…yeah...we'll be alright once we get rolling; he's supposed to be five miles by the time we're airborne, that's the idea". Shortly after takeoff the plane encountered wake turbulence, to which the first officer responded with strong aileron inputs. Immediately after the encounter the captain stated: "Little wake turbulence, huh?", to which the first officer replied "Yeah". After a few more seconds, a second wave of wake turbulence was encountered, to which the first officer reacted with strong rudder and aileron inputs. His aggressive action on the flight controls caused the plane to experience significant lateral oscillation, which the first officer erroneously attributed to wake turbulence. As a result, he continued his action on the flight controls, causing the plane to experience increasing side loads and resulting in the loss of the tail and the engines.. Throughout the accident flight, the first officer seemed to be convinced that wake turbulence would be encountered, and that some type of action may be needed. Records indicated that the first officer's preoccupation with wake turbulence was not limited to the accident flight, as he had showed strong reactions to wake turbulence in earlier occasions.

- United Airlines 173, Portland, OR, December 28th 1978 – As the aircraft approached the arrival airport, a problem arose with the landing gear extension. As the gear was lowered, the crew heard a loud "thump, thump," and the airplane yawed to the right. The only gear lights that came on were those indicating the nose gear was down and locked. The flight crew elected to assess the problem while in a holding pattern. However, the fuel level was not adequately monitored, and fuel starvation occurred, which caused the plane to crash before reaching the airport. About one hour elapsed between the time the problem with the gear emerged and the time of the crash. The flight engineer was monitoring the state of the fuel throughout the last segment of the flight and voiced concern to the captain. The flight engineer even stated the amount of fuel, which, considering the fuel burn rate, gives a clear estimate on the amount of time until the fuel would be depleted. However, the captain continued on a path that would keep them in the air longer than the fuel supply allowed. The NTSB determined that the probable cause of the accident was the failure of the captain to properly monitor the aircraft's fuel state and to properly respond to the low fuel state and the crewmember's advisories regarding fuel state (1979). This resulted in fuel exhaustion to all engines. The inattention resulted from preoccupation with a landing gear malfunction and preparations for a possible landing emergency (NTSB, 1979). The only cues being considered were those associated with the landing gear, despite the dire fuel situation. The captain was unable to successfully process the information regarding the fuel state because his attention resources were exhausted dealing with the landing gear problem.

- KLM 4508, Tenerife, March 27th 1977 – Numerous flights were diverted to Tenerife after the Las Palmas Airport closed because of a terrorist attack. The sudden increase in traffic caused congestion at Tenerife so that a KLM Boeing 747 was forced to wait two hours, while another plane, which blocked the taxiway, boarded passengers and refueled. The KLM flight was eventually allowed to move, but takeoff was initiated before a clearance had been issued. The plane struck another Boeing 747 that was taxing on the runway, resulting in the worst accident ever in the airline industry. During the takeoff roll, the KLM flight crew warned the captain that they might not have been cleared for takeoff and that another plane might have been taxing on the runway. However, the captain seemed to be strongly convinced that they had been cleared for takeoff and discarded the flight crew's comments.

- Northwest Airlines 6231, Thiells, NY, December 1st, 1974 – As the aircraft was climbing in icing conditions, the pitot tube became clogged by ice, so that the airspeed indicator started working as an altimeter, indicating increasing airspeed as the plane climbed. The flight crew failed to recognize the problem and instead believed, despite the constant power setting and the climb attitude, that the airspeed was in fact increasing. They believed that this increase was due to the low weight of the aircraft. Their erroneous interpretation lasted throughout the flight, until the plane buffeted, stalled, and entered a rapid descent. The flight crew apparently believed that the buffeting was a high speed phenomenon – Mach buffeting – rather than a stall buffeting and neglected the possibility of a stall despite the indication from the shaker stick. The flight crew relied exclusively on the air speed indicators and their related warning systems, ignoring other pertinent cues pointing to a different problem than the one originally assessed.

- Eastern Airlines 401, Everglades, FL, December 29th, 1972 – As Eastern 401 approached Miami International Airport and lowered the landing gear, the light that indicated that the nose landing gear has lowered and locked failed to illuminate. The crew chose to depart the airport airspace to the west to assess the problem. The auto- pilot was engaged, and they proceeded to evaluate the indicator light and the gear status. As the flight continued, the autopilot became disengaged and a slight descent initiated. Prolonged focus on the landing gear problem prevented the flight crew from monitoring altitude and the plane proceeded to descend, eventually impacting the ground. The NTSB found that the three flight crewmembers were preoccupied in an attempt to ascertain the position of the nose landing gear and therefore neglected monitoring the flight instruments (1973). Much like the crew involved in the accident in Portland in 1978, this crew was focused on the problem with the landing gear and did not sample other cues relating to the state of the aircraft. The flight crew did not even hear the altitude alert which sounded as the aircraft descended through 1, 750 feet m.s.1., an indication that their resources were entirely devoted to the landing gear.

Conclusions and Research Indications

As illustrated in the accidents presented above, cognitive tunneling likely played a role in several aircraft crashes. In all of the accidents discussed above, the pilot did not adequately perceive or evaluate all pertinent information necessary to successfully complete the flight because of filtering based on preexisting expectations, initial interpretations, or preoccupation with one aspect of the flight. The dilemma is evident; however, the solution is not so lucid. Prince et al. (1997) suggest three remedies that can be applied to overcoming the effects of stress in the cockpit: 1) redesign of task/environment, 2) selection of crew based on ability to withstand stressors, and 3) training, the most reasonable intervention. Prince et al.. suggest specific training techniques that appear promising including: integrating specific behavioral techniques designed to assist in dealing with stress, and providing crews the opportunity to practice newly acquired skills under condition of graded exposure to stressors (1997). Glyn (1997) suggested developing a comprehensive aircrew decision making seminar to include awareness training and incorporate pertinent research. In stead of presenting a specific formula for optimal decision making, a range of different decision types is presented along with the different processes used in making a good decision (Glyn, 1997). There are currently pilot training programs that do incorporate stress management and decision making training into their Crew Resource Management (CRM) training (Prince et al., 1997). However, alterations to these programs to include awareness training of the phenomena that can occur as a result of stress, such as cognitive tunneling, may prove beneficial. By exposing pilots to the theories and the research into the effects of stress on performance, and by illustrating these effects through previous accidents and occurrence in actual simulator training, pilots' susceptibility to it may decrease. Further research on pilot training with respect to stress and its effects is needed to better understand how to cope with this issue.

References

Chajut, E. & Algom, D. (2003). Selective attention improves under stress: Implications for theories of social cognition. *Journal of Personality and Social Psychology 85* (2), 231–248.

Comision de Accidentes e Incidentes de Aviacion Civil (1977). Report A-102/1977 y A-103/1977, Madrid, Spain.

Deitch, E.L. (2002). Learning to land: A qualitative examination of pre-flight and in-flight decision-making processes in expert and novice aviators. *Dissertation Abstracts International: Section A: Humanities and Social Sciences, 63(1-A),* 158.

Driskell, J.E., Salas, E., & Jonston, J. (1999). Does stress lead to a loss of team perspective? *Group Dynamics: Theory, Research Practice, 3* (4), 291-302.

Glyn, D. (1997). Decision making training for aircrew. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision making under stress*. Aldershot, England: Ashgate Publishing Company.

Guilkey, J.E. (1997). An investigation of aviation problem-solving skills as they relate to amount of total flight time. *Dissertation Abstracts international: Section B: The Sciences and Engineering, 58(5-B),* 2710.

Harris, R.L., Tole, J.R., Stephens, A.T., & Ephrath, A.R. (1982). Visual scanning behavior and pilot workload. *Aviation, Space, & Environmental Medicine, 53(11),* 1067-1072.

National Transportation Safety Board. (2004). *In-flight separation of vertical stabilizer American Airlines flight 587 Airbus Industries A300-605R, N14053, Belle Harbor, New York, November 12, 2001*. Washington D.C.: The National Transportation Safety Board.

National Transportation Safety Board. (1979). *United Airlines, Portland, Oregon, December 28th, 1978.* Washington D.C.: The National Transportation Safety Board.

National Transportation Safety Board. (1975). *Aircraft Accident Report: United Airlines, Thiells, New York, December 1st, 1974* Washington D.C.: The National Transportation Safety Board.

National Transportation Safety Board. (1973). *Eastern Airlines, Lockheed 1011, Everglades, FL, December 29th, 1972.* Washington D.C.: The National Transportation Safety Board.

Peres, M., Van De Moortele, P.F., Pierard, C., Lehericy, S., Le Bihan, D., & Guezennec, C.Y. (2000). Functional magnetic resonance imaging of mental strategy in a simulated aviation performance task. *Aviation, Space, & Environmental Medicine, 71(12),* 1281-1231.

Prince, C., Bowers, C.A., & Salas, E. (1997). Stress and crew performance: Challenges for aeronautical decision making training. In N. Johnston, N. McDonald, & R. Fuller (Eds.), *Aviation psychology in practice*. Aldershot, England: Ashgate Publishing Company.

Ross, L., & Lepper, M. R. (1980). The perseverance of beliefs: Empirical and normative considerations. In R. A. Schweder (Ed.), *New directions for methodology of behavioral science: Fallible judgment in behavioral research.* San Francisco: Jossey-Bass. (p. 98)

Wickens, C.D. & Flach, J.M. (1988). Information processing. In E. Weiner and D. Nagel (Eds.), *Human factors in aviation*. San Diego, CA: Academic Press.

Wickens, C.D., & Hollands, J.G. (2000). Engineering psychology and human performance. Upper Saddle River, NJ: Prentice Hall.

Wickens, C.D., Stokes, A., Barnett, B., & Hyman, F. (1993) The effects of stress on pilot judgment in a MIDIS simulator. In O. Svenson & A. J. Maule (Eds.), *Time pressure and stress in human judgment and decision making.* New York: Plenum Press.

COMMUNICATIONS BETWEEN TEAM MEMBERS OF DIFFERENT CULTURES AND NATIONALITIES ON INTERNATIONAL AIRLINE FLIGHT DECKS

Dr. Theodore N. Beneigh
Embry-Riddle Aeronautical University
Daytona Beach, FL

International flight operations became commonplace in the 1950s with the introduction of jet transport aircraft. The new jets had speeds that were twice as fast as the piston aircraft they were replacing, a range great enough to transit oceans nonstop, and a lower operating costs that made international travel affordable to many more people.

For the most part, most of the pilots flying these aircraft were natives from the airline's home country. As international operations expanded exponentially, many airlines had difficulty finding native-born pilots to fly their aircraft. The human resource departments of many airlines began to recruit new pilots globally. While most of these airlines had programs in place to teach rudimentary crew resource management procedures, the bulk of the training the pilots received concerned the technical operation of the aircraft and the takeoff, enroute, and arrival operations the pilots could expect during actual line operations. Very little training was given to the pilots in how to communicate effectively with people from different cultures. In addition, many pilots and air traffic controllers had difficulty clearly speaking and understanding English, which is the international language of aviation. This has had a negative impact on flight safety during international flight operations.

This presentation will show the results of the experimental method, which was selected to test five hypotheses:

1. Small group instruction techniques have no effect on improving authoritative/assertiveness interactions between pilots on culturally mixed flight decks.
2. Small group instruction techniques have no effect on improving the decision-making capabilities between pilots on culturally mixed flight decks.
3. Small group instruction techniques have no effect on improving trust between pilots on culturally mixed flight decks.
4. Small group instruction techniques have no effect on interpersonal relationships between pilots on culturally mixed flight decks.
5. Small group instruction techniques have no affect on improving the team atmosphere between pilots on culturally mixed flight decks.

The results of the experiment proved that training can improve the authoritative/assertiveness and team atmosphere characteristics of Asians and non-native English speaking pilots, and training can improve the interpersonal relationships and team atmosphere for Anglo and native English speaking pilots.

Proper administration of this training can lead to safer international flight operations.

Introduction

The deregulation of the United States airline industry by Congress in 1978 was the beginning of a revolution in the airline industry. Most aviation and consumer experts were of the incorrect belief that the affects of deregulation would be limited to the United States.

Before 1978, the United States airline industry was controlled by various agencies of the United States government (Taylor, 1964). These agencies set fares, determined schedule frequency, and determined which airline would serve which locations. The airlines were free to determine what type of aircraft they would use to fly the routes, and the service offered to the passengers on the ground and during flight. Since the airlines had no control concerning the fares and city pairs they served, competition between the airlines to fill their airplanes' seats created a level of service to the passengers served that would make their airline the most comfortable airline to fly.

October 26, 1958, was an historical day in international airline operations. Pan American World Airways began the first international non-stop jet service when it inaugurated flights between New York City and Paris (The Boeing Company). Before 1958, relatively slow propeller driven airplanes conducted nonstop, transoceanic service. A non-stop flight from London to New York required seventeen hours (Airline History). The few airlines that offered this service were mostly piloted by members from their home country (e.g. aircraft flown by British Overseas Airways Corporation were piloted by British pilots; aircraft flown by Pan American World Airways were flown by pilots from the United States). Nearly all flight decks on international airliners were multicultural. The Boeing 707 required seven hours and forty minutes (Official Airline Guide). Affordable, rapid, comfortable international air transportation was now available with the advent of the new, long-range jet transport.

The intent of deregulation was to place the airline industry in the United States into the realm of free-market competition. The initial result of deregulation was the airlines' reevaluation of the routes flown and cities served. In an effort to minimize the effects of competition, most carriers segregated their route structures. This segregation allowed one airline to provide the majority of the air service into a particular city, and thereby dominate the fares in the markets involving that city. Without significant competition, the airline could set its fares based on its perception of what the market would bear.

In the early 1970s, just before the United States airline deregulation, a select few air carriers offered the majority of international service offered worldwide. Pan American World Airways, Trans World Airlines, and British Overseas Airways Corportation flew most of this service. After airline deregulation in the United Sates, several major United States airlines began to realize the importance of international service for increased profitability and continued expansion. In addition, aircraft manufacturers began to manufacture aircraft with the range and cargo capacity that could make international routes very profitable.

Foreign air carriers also began to increase their international operations. Malaysia-Singapore Airlines segregated in 1972 to create the two large Pacific carriers of Malaysia Air and Singapore Airlines (Singapore Airlines). British European Airways merged with British Overseas Airways Corporation in 1975 to form British Airways (Airline History). In 1983, United Airlines began operations between the United States and Tokyo. In 1985, United Airlines acquired Pan American Airways' Pacific Division (United Airlines, Era 7). Several other United States flag carriers including Northwest, Delta, and American began setting up an international route structure.

For the most part, most of the pilots flying these aircraft were natives from the airline's home country. As international operations expanded exponentially, many airlines had difficulty finding native-born pilots to fly their aircraft. The human resource departments of many airlines began to recruit new pilots globally. While most of these airlines had programs in place to teach rudimentary crew resource management procedures, the bulk of the training the pilots received concerned the technical operation of the aircraft and the takeoff, enroute, and arrival operations the pilots could expect during actual line operations.

Many of these aircraft were being operated with two or more pilots with different nationalities and different cultures. Additionally, language was a problem. While the international language of aviation is English, many air traffic controllers in non-English speaking countries used their native language instead of English. Additionally, before the introduction of culturally mixed flight decks, verbal communication between the pilots was usually conducted in their native language.

Flight operations with culturally mixed flight decks have created a plethora of problems, including language, a conflict of cultural norms, and the role of command/subordination on the flight deck.

Methodology

The experimental method will be selected to test the five hypotheses. Research done by Hanssen, Stayton, and Wlaka (1992) concerning multi-cultural considerations for space station training, and operational issues created by cultural differences that can pose potential safety problems (Helmreich, 2000) justify the need for this experiment. Using that data, the problems identified in the KLM/Pan American collision (National Transportation Safety Board, 1978) and the Flying Tiger 66 accident (Continental Airlines, 1989), current crew resource management practices (United Airlines, 1995), and my operational experience as a pilot teaching crew resource management to pilots from different cultures, the following cultural relationships will be measured:

1. Authoritativeness and assertiveness
2. Decision-making
3. Trust
4. Interpersonal relationships
5. Team atmosphere

Since industry implementation of the training will involve training culturally mixed and monocultural crews in training centers located throughout the world, two groups will be used to test the hypothesis. The first group will involve the selection of an equal number of Anglo and Asian participants. While all the participants will speak English, one-half of the participants will be native English speakers, and the other half will be non-native English speakers. An expert in crew resource management training who possesses expertise in cross-cultural training (Hanssen et al., 1992) will administer the training. The second group will involve the performance of the experiment in an Asian country. The participants will be monocultural, and the trainer will be an expert in crew resource management training who possesses an archetypal expertise in cross-cultural training (Hanssen et al., 1992).

All participants will be given a pretest. The pretest will consist of two scenarios that are representative of situations international flight crews can experience (United Airlines, 1995). Each scenario will have 10 questions. The answers to these questions will be indicative of how well the participants will communicate with their fellow crewmembers by measuring their responses in the five cultural being measured in the hypotheses (Hanssen et al., 1992).

At the conclusion of the pretest, the training will begin. Based on my experience teaching crew resource management for 15 years, and the principles identified by Hanssen et al. (1992), the training will consist of training in:

- Communications principles as they relate to crew resource management and flight crews.
- A guided discussion on barriers to communication and suggested solutions.
- A discussion of three replayed scenarios viewed by the group on a television.
- The participants will then be grouped into pairs, mixing Anglo with Asian pilots, into two-men flight crews.

Each flight crew will fly an identical flight training scenario, which will involve an in-flight emergency. At the conclusion of the flight training scenario, each participant will be given a post test. The post test will consist of two scenarios, each of which will have 10 questions. The questions will measure the same five cultural relationships that will be measured in the pretest.

This type of training is representative of the type of crew resource management training given by international airlines (United Airlines, 1995). The data will be collected from Anglo and Asian pilots, and from native English and non-native English speaking pilots. The answers to the pretest and post test questions will be given a numerical value. A value of one will be assigned to a strongly agree response; two for agree, three for uncertain, four for disagree, and five for agree. A higher number will indicate a more desirable position to effectively communicate.

A test for normality will be performed. If normality exists, a parametric test, such as a T test, will be applied. If normality does not exist, a non-parametric test, such as the Mann-Whitney U test, will be used. A non-parametric Sign test will be used to measure the difference in response between the pretests and the post tests to individual from the same culture. The statistical analysis of this data will determine if the training was effective in improving communications between team members of different cultures and nationalities on an international airline flight deck.

Experimental Results

Group One
Comparison of the Anglo and Asian Cultures

The purpose of this paper is to test the effectiveness of training to improve communications between team members of different cultures and nationalities on international airline flight decks. The communication skills were broken down into five areas: authoritative/assertiveness, decision-making, trust, relationships, and team atmosphere.

Prior to the beginning of the training, a comparison was made to assess the differences between the Asian and Anglos cultures by comparing their answers to the pretest questions.

Concerning authoritative/assertiveness, the Anglos and Asians were identical. This can be accounted for considering the common specific training, and the behaviors the pilots expected from the fellow crewmembers and air traffic control, that are given pilots worldwide flying transport category jet aircraft in international operations.

Decision-making had similar results. While both cultures were similar, the Asians had a slightly higher mean score. This is most likely accounted for by the higher Power/Distance Index (PDI) characteristic of Asian cultures when compared to Anglo cultures (Hofstede, 1991).

Concerning trust, the Asians had a higher value, in both the range of scores and the mean, in the pretest when compared to the Anglos. This can be explained by the higher Uncertainty Avoidance Index (UAI) characteristic of Asian cultures when compared to Anglo cultures (Hofstede, 1991). Cultures with a high UAI value tend to be set in their ways, skeptical of new thought and ideas. Cultures with a low UAI value are more open to new ideas, are more creative, and more willing to take chances.

The relationships between the crewmembers are very important for the team to function effectively. The Anglos and Asians were identical. This also can be accounted for considering the common specific training, that are given pilots worldwide flying transport category jet aircraft in international operations.

In the characteristic of team atmosphere, the Anglos had a higher range and mean than the Asians when comparing the pretests. The Individualism/Collective Index (IDV) characteristic of the Anglo culture is higher than those of the Asian cultures (Hofstede, 1991). A high IDV value represents a culture that places a reward for individual initiative, emphasizing the importance of individual thought and creativity. A low IDV value reflects a culture more comfortable working in groups. These results of the pretest contradict what can be expected by the IDV values. It would be expected that the lower IDV groups would have a higher team atmosphere. It may be possible that the strict training and importance of teamwork has a greater affect on the Anglos, causing them to have a higher score in team atmosphere.

Analysis of the Results of the Experiment

The results of the experiment had different results, depending on the hypothesis being considered.

Hypothesis 1: Small group instruction techniques have no effect on improving authoritative/ assertiveness interactions between pilots on culturally mixed light decks. By comparing the pretest administered to the Anglos before the training with the post test administered after the training, this hypothesis is supported. The Sign Test value of .302 indicated no difference between Anglo pretest and post test scores. For the Asian group, the hypothesis is rejected. The Sign Test value of .001 reflects an improvement in the authoritative/assertiveness characteristic. These results are reflective of the nature of the Asian culture. Asians cultures have a high PDI value. This indicates an acceptance of hierarchy as an important element of human behavior. Hence, proper training is more likely to affect a culture with a high PDI value in authoritative/assertiveness than a culture that places a lower emphasis on these values, such as the low PDI Anglos cultures.

Hypothesis 2: Small group instruction techniques have no effect on improving the decision-making capabilities between pilots on culturally mixed flight decks. By comparing the pretest administered to the Anglos before the training with the post test administered after the training, this hypothesis is supported. The Sign Test value of .210 indicated no difference between Anglo pretest and post test scores. By comparing the pretest administered to the Asians before the training with the post test administered after the training, this hypothesis is also supported. The Sign Test value of .077 indicated no difference between Asian pretest and post test scores. This can be explained by the fact that the training model was ineffective. Previous attempts at teaching decision-making have failed. Different cognitive skills are involved with teaching decision-making, with each individual responding to different cognitive skills, and past efforts at training general purpose cognitive skills have met with failure (Bransford, Arbitman-Smith, Stein & Vye, 1985).

Hypothesis 3: Small group instruction techniques have no effect on improving trust between pilots on culturally mixed flight decks. By comparing the pretest administered to the Anglos before the training with the post test administered after the training, this hypothesis is supported. The Sign Test value of .210 indicated no difference between Anglo pretest and post test scores. By comparing the pretest administered to the Asians before the training with the post test administered after the training, this hypothesis is also supported. The Sign Test value of .581 indicated no difference between Asian pretest and post test scores. These results can be explained by the fact that one cultural characteristic that is typical of all pilots is that they are highly individualistic in nature (Weiner et al., 1993, p. 68). This characteristic makes them wary of changing their trust in other pilots, thereby making it difficult to increase their trust in other pilots.

Hypothesis 4: Small group instruction techniques have no effect on interpersonal relationships between pilots on culturally mixed flight decks. By comparing the pretest administered to the Anglos before the training with the post test administered after the training, this hypothesis is rejected. The Sign Test value of .007 reflects an improvement in the interpersonal relationships between pilots. These results are reflective of the nature of most Anglo cultures. Anglo cultures tend to have a high IDV value. This represents a culture that places a reward for individual initiative and favors individual initiative over group activity (Hofstede, 1991). Hence, proper training is likely to improve an individual with high IDV values. By comparing the pretest administered to the Asians before the training with the post test administered after the training, this hypothesis is supported. The Sign Test value of 1.00 indicates no difference between pretest and post test scores. Most Asian cultures have a low IDV value (Hofstede, 1991). Individuals from these cultures prefer group activity to individual activity, so are more likely to have good interpersonal relationships before the training, making this characteristic more difficult to improve.

Hypothesis 5: Small group instruction techniques have no affect on improving the team atmosphere between pilots on culturally mixed flight decks. Most Anglo cultures favor individualism, and most Asian cultures favor collectivism (Gudykunst, 1994). It could be inferred that Anglo cultures would be more likely to improve in the characteristic of team atmosphere that Asians. Such is not the case. By comparing the pretest administered to both Anglos and Asians before the training with the post test administered after the training, this hypothesis is rejected for both the Anglos and the Asians. The Sign Test for the Anglos was 0.000, and for the Asians the Sign Test was .013. This indicates a successful training program in improving team atmosphere for both cultures.

Group Two

Analysis of the Results of the Experiment

Hypothesis 1: Small group instruction techniques have no effect on improving authoritative/assertiveness interactions between pilots on culturally mixed light decks. By comparing the pretest administered to the Asians before the training with the post test administered after the training, this hypothesis is rejected. The Sign Test value of .031 reflects an improvement in the authoritative/ assertiveness characteristic. These results are reflective of the nature of the Asian culture. Asians cultures have a high PDI value. This indicates an acceptance of hierarchy as an important element of human behavior. Hence, proper training is more likely to affect a culture with a high PDI value in authoritative/assertiveness than a culture that places a lower emphasis on these values, such as the low PDI Anglos cultures.

Hypothesis 2: Small group instruction techniques have no effect on improving the decision-making capabilities between pilots on culturally mixed flight decks. By comparing the pretest administered to the Asians before the training with the post test administered after the training, this hypothesis is supported. The Sign Test value of .375 indicated no difference between Asian pretest and post test scores. This can be explained by the fact that the training model was ineffective. Previous attempts at teaching decision-making have failed. Different cognitive skills are involved with teaching decision-making, with each individual responding to different cognitive skills, and past efforts at training general purpose cognitive skills have met with failure (Bransford, Arbitman-Smith, Stein & Vye, 1985).

Hypothesis 3: Small group instruction techniques have no effect on improving trust between pilots on culturally mixed flight decks. By comparing the pretest administered to the Asians before the training with the post test administered after the training, this hypothesis is supported. The Sign Test value of .375 indicated no difference between Asian pretest and post test scores. These results can be explained by the fact that one cultural characteristic that is typical of all pilots is that they are highly individualistic in nature (Weiner et al., 1993, p. 68). This characteristic makes them wary of changing their trust in other pilots, thereby making it difficult to increase their trust in other pilots.

Hypothesis 4: Small group instruction techniques have no effect on interpersonal relationships between pilots on culturally mixed flight decks. By comparing the pretest administered to the Asians before the training with the post test administered after the training, this hypothesis is supported. The Sign Test value of .063 indicates no difference between pretest and post test scores. Most Asian cultures have a low IDV value (Hofstede, 1991). Individuals from these cultures prefer group activity to individual activity, so are more likely to have good interpersonal relationships before the training, making this characteristic more difficult to improve.

Hypothesis 5: Small group instruction techniques have no affect on improving the team atmosphere between pilots on culturally mixed flight decks. By comparing the pretest administered to the Asians before the training with the post test administered after the training, this hypothesis is supported. The Sign Test value of .219 indicates no difference between pretest and post test scores. Again, most Asian cultures have a low IDV value (Hofstede, 1991). Individuals from these cultures prefer group activity to individual activity, so are more likely to have good team atmosphere prior to beginning the training, making an improvement in this characteristic more difficult.

Conclusions

In all experimental scenarios, all Anglos were native English speakers, and all Asians were non-native English speakers. Hence, the following conclusions can be made:

Hypothesis 1: Small group instruction techniques have no effect on improving authoritative/ assertiveness interactions between pilots on culturally mixed light decks. Training can improve this characteristic for Asians and non-native English speakers. Training cannot improve this characteristic for Anglos and native-English speakers.

Hypothesis 2: Small group instruction techniques have no effect on improving the decision-making capabilities between pilots on culturally mixed flight decks. Training cannot improve this characteristic for Asians, non-native English speakers, Anglos, or native English speakers.

Hypothesis 3: Small group instruction techniques have no effect on improving trust between pilots on culturally mixed flight decks. Training cannot improve this characteristic for Asians, non-native English speakers, Anglos, or native English speakers.

Hypothesis 4: Small group instruction techniques have no effect on interpersonal relationships between pilots on culturally mixed flight decks. Training cannot improve this characteristic for Asians and non-native English speakers. Training can improve this characteristic for Anglos and native-English speakers.

Hypothesis 5: Small group instruction techniques have no affect on improving the team atmosphere between pilots on culturally mixed flight decks. In a culturally mixed training environment, training can improve this characteristic for Asians, non-native English speakers, Anglos, and native English speakers. In a monocultural training environment, training cannot improve this characteristic for Asians or non-native English speakers. This can be explained by the intercultural communication that occurs in the training environments where the cultures are mixed, and the lack in intercultural communication that occurs in a monocultural training environment.

The only advantage to the Asian pilots in comparing the results of the training to the pilots of mixed cultures, Group 1, to the training of monocultural pilots, Group 2, is the improvement in team atmosphere to the training administered to the culturally mixed group. Two factors may have affected this outcome. Since an experienced airline instructor did the monocultural training to pilots from that one, specific airline, there may have been some conflicts of training with established company policies. This further enhances the need for corporate organization and philosophy to be supportive of goals of intercultural training. Secondly, the pilots from Group 1 were all very experienced line pilots with years of operational experience in multi-pilot aircraft. This may make improving team atmosphere more difficult since those pilots are more “set in their ways” of doing things.

It can be concluded that training in improving communications between team members of different cultures and nationalities on international airline flight decks is successful in improving relationships and team atmosphere in Anglo, native English speaking cultures, and in improving authoritative/assertiveness and team atmosphere in Asian, non-native English speaking cultures.

It can be concluded that training in improving communications between team members of different cultures and nationalities on international airline flight decks is not successful in improving authoritative/assertiveness, decision-making skills, and trust in Anglo, native English speaking cultures, and is not successful in improving decision-making skills, trust, and relationships in Asian, non-native English speaking cultures.

References

Airline history – airlines by index. (n.d.) *B.O.A.C. United Kingdom – G.* Retrieved February 17, 2004, from http://airlines.afriqonline.com/airlines/560.htm

The Boeing Company. (n.d.) *The 707 family*. Retrieved February 17, 2004, from http://www.boeing.com/commercial/707family/

Bransford, J.D., Arbitman-Smith, R., Stein, B.S., & Vye, N.J. (1985). Improving thinking and learning skills: An analysis of three approaches. In J. W. Segal, S. F. Chipman, & R. Glaser (Eds.). *Thinking and learning skills,* (Vol. 1, pp. 133-208). Hillsdale, NJ: Erlbaum.

Gudykunst, W. B. (1994). *Bridging differences, effective intergroup communication.* (2nd ed.). London: Sage Publications.

Hanssen, V., Stayton, W., & Wlaka, M. (1992*). Multi-cultural considerations for space station training and operations*. AIAA Space Programs and Technologies Conference, March 24-27, 1992. Huntsville, AL: Authors.

Helmreich, R.L. (2000). Culture and error in space: Implications from analog environments. *Aviation, Space, and Environmental Medicine*, 71(9-II), 133-139.

Hofstede, G. (1991). *Cultures and organizations: software of the mind.* Berkshire, U.K.: McGraw-Hill.

National Transportation Safety Board. (1978) *Aeronautical collision Boeing 747 PH-BUF and KLM Boeing 747.* Washington, D.C.: U.S. Government Printing Office.

Official airline guide North American executive flight guide. (August, 2003) 34 (8), p. 449.

Singapore Airlines. (n.d.). *About SIA SIA story.* Retrieved February 24, 2004, from http://www.singaporeair.com/sg

Taylor, F. J. (1964). *High horizons.* New York: McGraw-Hill.

United Airlines. (n.d.) *Era 7: 1970-1989.* Retrieved February 24, 2004, from http://www.united.com/page/middlepage/0,6823,2292,00.html

United Airlines. (1995). *Flight operations manual*. Englewood, CO: Jeppesen Sanderson, Inc.

Wiener, E. L., Kanki, B. G., & Helmreich, R. L. (1993). *Cockpit resource management.* San Diego, CA: Academic Press.

THE EFFECT OF TERRAIN-DEPICTING PRIMARY-FLIGHT-DISPLAY BACKGROUNDS AND GUIDANCE CUES ON PILOT RECOVERIES FROM UNKNOWN ATTITUDES

Dennis B. Beringer & **Jerry D. Ball**
The FAA Civil Aerospace Medical Institute
Oklahoma City, Oklahoma

Kelly Brennan & **Sitafa Taite**
The University of Oklahoma
Norman, Oklahoma

A study was conducted to evaluate the effects of primary flight display (PFD) terrain depictions on pilots' performance of recoveries from unknown attitudes. Forty pilots participated in the study, each group of eight using a different display format. The five conditions consisted of combinations of terrain depiction (none, full-color terrain, brown terrain) and guidance indications (pitch and roll arrows). Participants flew baseline trials in the Advanced General Aviation Research Simulator using a common electronic attitude indicator and then performed recoveries from unknown attitudes (UARs) using one of the PFD formats. Performance measures included initial response time, total recovery time, primary reversals, and secondary reversals. No significant effects of the primary independent variables were found on any of the performance measures. Posttest interviews indicated the participants preferred the directional-arrow indicators and had no preference for or against the presence of terrain depictions during UARs, focusing primarily on the zero-pitch line as a reference. It was concluded that the specific terrain representations examined did not pose a hazard to the identification of and recovery from unknown attitudes as long as a zero-pitch line of sufficient discriminability (contrast and size) to all backgrounds was present.

Background

Electronic Flight Instrumentation Systems (EFIS) are becoming more available daily, and a major component of this type of system is the Primary Flight Display (PFD). While PFDs initially depicted attitude and flight-guidance information, they evolved to include forward-looking perspective-views of both guidance information (Beringer, 2000) and of the outside world (Wickens, Haskell, & Hart, 1989; Alter, Barrows, Jennings, & Powell, 2000), often generated from terrain databases. This type of display is presently appearing in systems submitted for certification in general aviation (GA) aircraft, and a number of questions have been raised regarding the effects of various design features on different aspects of pilot performance. In lieu of empirical data on the effects of manipulations of specific design parameters, certifiers have had to rely upon general guidelines. This has sometimes resulted in the adoption of very conservative criteria for the certification and use of these particular displays.

Some data relevant to the GA environment have become available that may be useful for determining what the allowable range of variation in design parameters can be. The parameters that are of present interest include: size of the display, angular representation of the outside world (field of view), display resolution, terrain-feature resolution, use of color, style of terrain representation, definition of display clutter, and effects of the above on the performance of both routine and non-routine flight tasks.

A series of studies was performed at the NASA Langley Research Center examining the use of various terrain representations and pilot preferences for various fields of view and styles of depiction (Prinzel et. al., 2003; Arthur, Prinzel, Kramer, Parrish, & Bailey, 2004). Some agreement was found with previous studies concerning preference for field of view (30 degrees), and some assessment was made of pilot navigation performance and basic precision maneuvers, concluding that fewer errors were committed and terrain awareness was enhanced with the displays. One issue that was not addressed, however, was the recovery from unknown or unusual attitudes. This specific concern was addressed in one certification process by requiring that the terrain depiction be removed from the PFD when the aircraft exceeded certain pitch or roll criteria because of a concern that the presence of the terrain might cause confusion or somehow interfere with a successful recovery. However, there were no empirical data to indicate what role, positive or negative, the terrain depiction might play in the recoveries.

Thus, a study was conducted to examine how terrain depiction might either impede or enhance recoveries from unknown attitudes, including the display content (type of terrain; flat, mountainous) at the time of the recovery as well as the possible ameliorating effect of providing recovery guidance arrows (Gershzohn, 2001). Questions of specific interest were: (1) would pilots recover to the terrain horizon rather than the zero-pitch line if the two were different, as would be seen in mountainous terrain; (2) if this behavior were observed, could it be ameliorated by positive guidance cues; and (3) would the coloration of the terrain presentation affect performance?

Method

Experimental Display Formats

The five display formats consisted of combinations of terrain depiction (none, full-color terrain, brown terrain) and guidance indications (pitch and roll arrows).

Baseline ADI. The no-terrain display consisted of a traditional attitude indicator (blue sky, brown ground) with airspeed, altitude and vertical speed presented in tape format along the left and right edges of the display with a compass card at the bottom of the display (Figure 1).

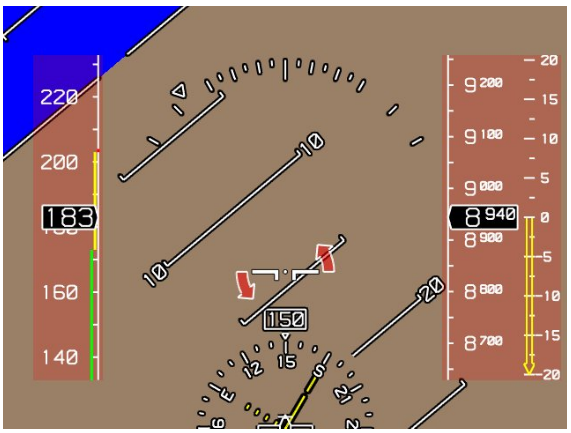


Figure 1. *EADI with roll-recovery arrow shown.*

Guidance Arrows. The second display was identical to the first but had guidance arrows for pitch and roll recovery. Pitch arrows were linear (Figure 2) and appeared when the aircraft attitude was greater than 13 degrees up or down and disappeared when the aircraft was within 5 degrees of zero pitch, pointing from the aircraft symbol to the horizon. Roll arrows (Figure 1) were curvilinear (arc form) and appeared when the aircraft exceeded 25 degrees of bank and disappeared when the aircraft was within 10 degrees of zero bank, pointing from the plane of the wings to the horizon line. For pitch-down attitudes, the roll-command arrow took precedence over the pitch-command arrow. For pitch-up attitudes, the priority was reversed.

Full-color terrain. The third display was similar to the first except that the brown portion of the display was replaced with photo-realistic terrain (full-color; shown in both Figures 2 and 3). The terrain was generated using variable-sized polygons that had photo-realistic texture applied to them to create the out-the-window scene. This is somewhat different from the terrain-creation methods used by other terrain-depicting displays, where equal-sized polygons, or even squares, are used to create the terrain skin and a more generic type of texture is applied. The fourth display was the same as the third display, but it included the guidance arrows.

Figure 2. *PFD with pitch-recovery arrow shown.*

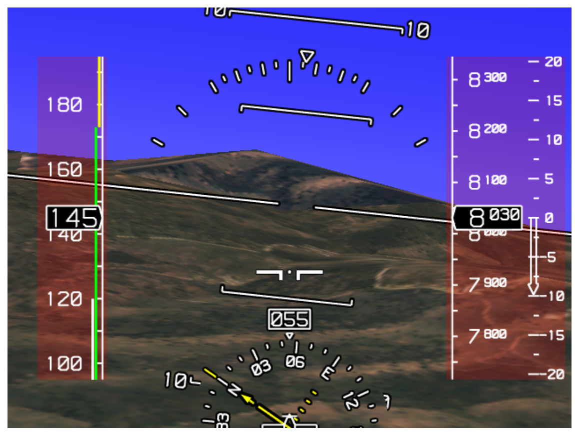


Figure 3. *PFD full-color terrain depiction with mountain in view.*

Brown terrain. The final display was similar to the first, but the "ground" portion of the display was replaced with brown (polygon-based) terrain imagery (Figure 4). The variable-sized-polygon structure im-

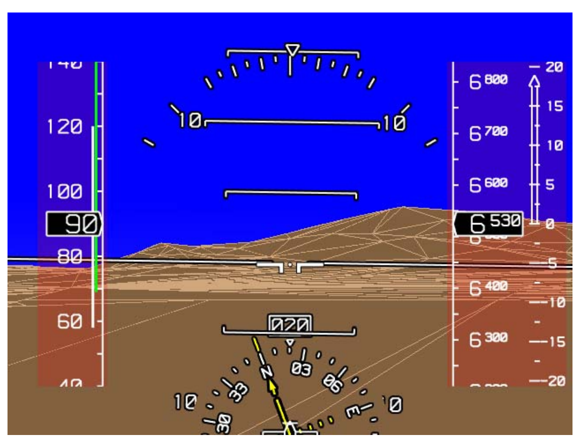


Figure 4. *PFD brown-only terrain depiction with mountain in view.*

parted more apparent texture to this uniform-brown depiction than one sees in brown-only depictions using a uniformly sized polygon/square as the basis for terrain-contour construction. Figures 3 and 4 show similar views of a mountain in the full-color (Figure 3) and the brown-only (Figure 4) modes for comparison.

Horizon line. The horizon line was constructed such that it would have high contrast against the vast majority of possible backgrounds. This is not normally an issue with traditional head-down attitude direction indicators (ADIs), as the horizon on these displays is represented as the boundary between differently colored filled areas, often with a line of a different color between them. It is also possible to use a single-color line (as long as it conforms to MIL-STD-1787C, 5.1.2.1; Horizon reference; the standard does not deal specifically with terrain-depicting PFDs, nor does the soon-to-be-released SAE Aerospace Recommended Practice document on perspective displays deal specifically with this horizon-line issue) in terrain-depicting displays where the ground and sky representations are of known uniform colors (i.e., the Chelton display uses a uniformly brown ground and blue sky).

However, displays expected to portray a realistically colored terrain representation or an enhanced depiction having multiple, albeit unrealistic, hues require a horizon line having components (bands) that will contrast against many hues. To this end, a horizon line was employed consisting of three two-pixel bands alternating black-white-black. This was consistent with horizon lines used in other full-color terrain display experiments and with recommendations made to a certification applicant who was submitting a colored-terrain PFD for consideration.

The original display was created at a resolution of 640 by 480 pixels but presented on a 1280 by 1024 flat-panel display in the cockpit using 800 by 600 pixel resolution inset in the upper right portion of the display. This produced a PFD image approximately 7.5 inches wide by approximately 5.6 inches tall (a 9.38 diagonal) and increased the apparent horizon-line thickness from 6 pixels to about 8 pixels. Seen from the pilot's viewing distance of 26 inches, the active display subtended 16.4 degrees horizontally and 12.3 degrees vertically, with the three-banded horizon line subtending approximately 9.85 minutes of arc vertically (each band about 3.3 minutes of arc).

Experimental Design

A two-factor crossed design was employed, with terrain background (full-color; present or absent) and guidance arrows (present or absent) as the *independent variables*. The supplemental condition, brown-only terrain, was added after contribution of guidance arrows had been assessed. Dependent variables included initial response time (IRT; time to first control input), total recovery time (TRT), primary control-input reversals (first response in wrong direction), and secondary control-input reversals (subsequent response in wrong direction).

Two *sampling variables* were added to obtain more representative data from across a wider range of display indications. *Terrain depiction at roll-out* was planned using lead headings based upon expected roll-out times (obtained in pretest) and presented terrain either (1) higher than the zero-pitch reference line (mountainous background) or (2) terrain lower than the zero-pitch reference line (level terrain). *Attitude at recovery onset* was also varied so that trials included combinations of pitch (+20, 0, and –15 degrees) and bank (60 degrees left, 0, 60 degrees right) excepting, of course, the zero-zero condition.

Three supplemental trials were also added for approximately the last 7 pilots in each group. These trials included a near-mountains trial (terrain horizon significantly above zero-pitch line), an inverted trial (by sponsor request), and a 40-degree displayed field-of-view trial (to assess whether previously expressed pilot preferences for a wider displayed field of view was linked with any improvement in performance with a wider field).

Equipment and Participants

Data were collected using the Advanced General Aviation Research Simulator (AGARS) in the CAMI Human Factors Research Laboratory. The simulator was configured to represent a Piper Malibu; the participants all flew in the left seat. The PFD was represented on a flat-panel, high-resolution LCD mounted on the instrument panel directly in front of the participant. The PFD was presented at the size of an approximately 7-inch diagonal measurement within a larger hardware-display area, and the image showed approximately 30 horizontal degrees of the outside world.

The display layout was similar in many respects to one already certified for GA use. The experimenter-pilot (EP) flew from the right seat with a repeater display of the PFD mounted atop the glare shield. The out-the-window view represented a hard-IFR situation with no environmental visual cues visible in the uniformly gray fields. Performance data were recorded digitally, with supplemental audio and visual data recorded on DVD from two video sources (cockpit-wide view and PFD inset) and all audio sources (participant, EP, data-collection experimenter).

Participants were 40 GA pilots (38 male, 2 female) recruited from the local community, 8 assigned to each of the 5 display conditions. Age and overall flight hours were balanced across groups as participants entered the experiment (not assigned a priori from a known sample). Ages ranged from 19 to 57 years. All were at minimum certified as Private Pilot, while many were instrument-rated and a number were flight instructors; initial license year ranged from 1972 to 2004. Each group had a similar distribution of pilot categories and hours of experience represented, with total pilotage time (as PIC in VMC) ranging from 11 to 11,700 hours. Total flight times ranged from 50 to 13,000 hours.

Procedures/tasks

After completing the informed consent form and filling out a brief pilot experience questionnaire, participants were briefed concerning the display they would be using and instructed that recoveries would be from unknown attitudes. Their task was to recover to a zero-pitch, zero-bank attitude, regardless of altitude or airspeed, as the EP would configure the aircraft such that performance was usually within the operating envelope (primary interest was in participant ability to interpret the display and determine when a level attitude had been restored). They were then ushered into the AGARS, where they were further familiarized with the display and with the simulator. They then donned a headset and a visor so that direct vision of the display would be obscured when they were in the head-down preparatory position for the recovery.

Each pilot then took off from Albuquerque (ABQ) and climbed out to the north into IFR conditions. All pilots performed 8 warm-up (baseline) recovery maneuvers, using the basic electronic attitude-direction indicator (EADI) on the PFD, to familiarize them with the performance of the AGARS and with the dynamic functioning of the PFD. Each trial began with the participant in the head-down position and hands off of the controls. The EP then placed the simulator into the required attitude and heading for that trial, using predetermined airspeed, altitude, and heading criteria that had been rehearsed (the same EP performed all unknown-attitude entries for all participants). The EP gave a preparatory “Ready” about two seconds before handing over the controls, “and” about one second before, and “Go!” at the transfer of controls to the participant. After completing the warm-up trial, the participant flew the simulator back to ABQ and performed a full-stop landing. At this time, the display format was changed and the procedure repeated.

Experimental trials consisted of 16 recovery maneuvers (defined by combinations of the sampling variables described earlier), using the PFD that was assigned to the participant. Two different orders of the combinations of sampling variables (attitude at onset and terrain seen at roll-out) were used and balanced across the groups. Accordingly, half of the headings were selected to end the recovery facing mountainous terrain higher than the aircraft altitude and half were selected to end the recovery facing terrain lower than aircraft attitude. Pilot recovery times and initial response times were recorded for each trial. A recovery was considered complete when the aircraft reached ±2.5 degrees of pitch and ±5.0 degrees of bank and was able to maintain those values for 3 seconds, although trials were generally allowed to continue for a few seconds after these criteria had been reached to guarantee stability in the recovery.

The supplemental trials described earlier in the Methods section were added to the end of the session in the order of (1) near-mountains trial, (2) inverted trial (the nose slightly above the horizon and a bank angle of approximately 165 degrees), and (3) expanded FOV trial. The participant then flew the simulator back to ABQ for a full-stop landing. Participants completed a posttest set of questionnaires regarding their subjective assessment of the displays (one was also administered after the warm-up trials), went through a posttest interview, and provided both solicited and unsolicited responses/opinions.

Results

Group Equivalence

Demographic variables. Groups were compared both on the basis of the distributions of experience (hours), categories of license/ratings, and age. Mean age by group ranged from 26 to 28 years of age with no significant differences between groups. The distributions of hours of experience and licensing/rating categories were also similar enough between groups that any differences found in performance were unlikely to be a result of those variables.

Baseline performance. Analysis of recovery times for the baseline trials showed that the groups initially differed in their performance but were performing equivalently (no significant differences) by the last two trials (see Figure 5). This finding suggests that all groups had attained a roughly equivalent level of performance prior to entering the experimental trials.

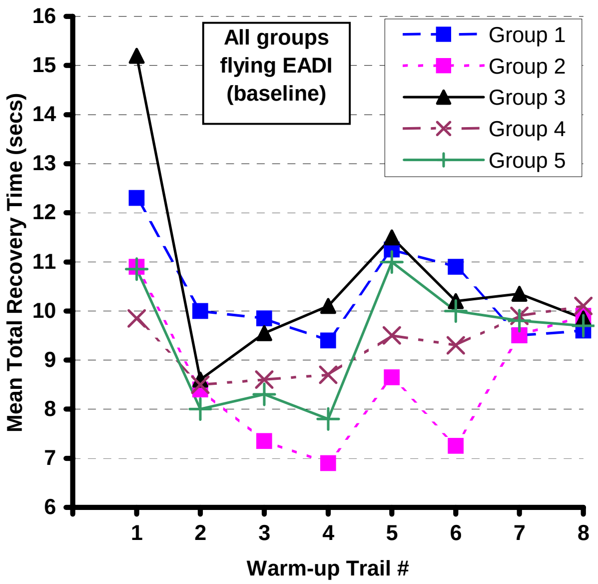


Figure 5. *Mean recovery time by group and serial trial for baseline warm-up using the basic electronic attitude direction indicator (EADI).*

Performance Variables

Recovery times. Multivariate Analysis of Variance indicated there were no significant differences between the display configurations for either (IRT, TRT) of the response-time variables. Figure 6 presents mean TRTs by maneuver and display format.

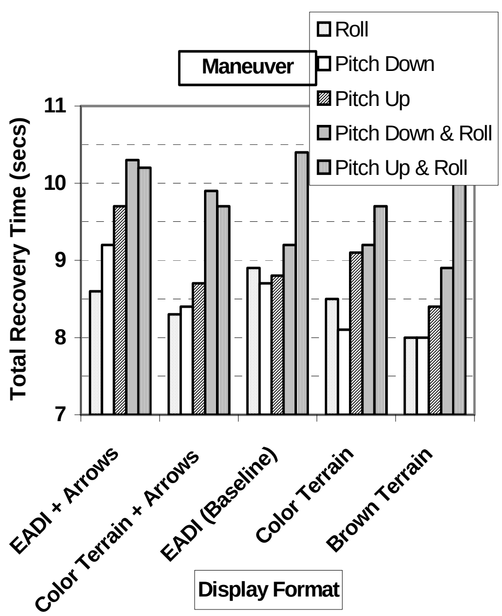


Figure 6. *Mean TRT (seconds) by maneuver and display format.*

To illustrate times actually required to complete a recovery, pitch-roll TRTs averaged around 10 seconds, whereas roll-only recoveries averaged about 8.5 seconds. Pitch-only recoveries averaged approximately 8.6 to 9.0 seconds. Univariate analyses were conducted to determine if type of maneuver resulted in any significant differences between display types. Again, no significant differences were found between displays and type of maneuver for either of the response-time measures. (Means by maneuver and display format are presented in Figure 6.)

Control reversals. Examination of control reversals, defined as movements in the opposite direction of that required for the recovery, indicated that were only three clearly identifiable primary control reversals in the nearly 800 trials. There were no secondary reversals (initial response in correct direction; subsequent control movement opposite to input required). Recovery times for the three reversals were not notably different from those of other trials. Thus, reversals did not appear to be a factor, regardless of the format of display used.

Supplemental trials. Analyses were conducted for performance variables on each of the three supplemental trials. No significant differences were found for the 40-degree FOV trials, the inverted trials, or the near-mountains trials. Only one of the participants showed any indication of holding the nose of the aircraft above the zero-pitch line in the near-mountain trial rather than completing the recovery.

Questionnaires and Posttest Interviews

Pilots indicated, when interviewed, that they were focusing their attention on the relatively prominent zero-pitch line, and did not regard the terrain depictions as significant contributors to their recovery task. The directional-guidance arrows produced a positive qualitative response from the participants, although there was no apparent performance difference. Participants also expressed a relatively uniform preference for the terrain-depicting displays in general. A few individuals expressed a preference for the 40-degree FOV, stating that it allowed them to "see more." The one individual who had kept the nose of the simulator slightly higher than zero pitch for the near-mountain trial clarified, in the posttest interview, that he had been concerned about the mountain and had kept the nose a little high in preparation for a possible climb over the mountain, having no indeterminacy about the zero-pitch line location.

Summary and Conclusions

It appears, for this specific task, that the presence of a zero-pitch line of the contrasting components specified (white with black borders) and of the thickness and extent specified (9 minutes of visual arc and running the entire width of the display area) allows pilots to adequately discern the zero-pitch reference from other features on the display and to perform recoveries from unknown attitudes without regard to the specific format of perspective terrain display used. It also appears that the directional-guidance arrows, despite being positively received by the participants and having been demonstrated to be useful in a previous experiment, did not have an appreciable effect on recovery times. The frequency of occurrence of reversals was too low to allow any conclusion to be drawn about the possible effectiveness of guidance arrows in that regard.

Given the previous findings (indicating enhanced terrain awareness attributable to terrain depictions), combined with the lack of detrimental effects found in this study relative to recoveries from unknown attitudes, there would appear to be few significant obstacles to the implementation of this type of PFD for general aviation use. Caveats to be observed, however, would be that (1) similarly constructed terrain depictions are used, (2) the zero-pitch line is clearly differentiable from the terrain and sky depictions regardless of the type of background and (3) that the direction of off-display pitch-line locations are clearly indicated.

Acknowledgments

The authors thank Howard Harris for contacting and scheduling the pilot participants in the study and Barry Runnels for his assistance with the operation of the AGARS. The OU authors, Kelly Brennan and Sitafa Taite, participated in the study at part of their Capstone project in the School of Industrial Engineering, The University of Oklahoma, and we thank the faculty advisor and program organizers for providing the opportunity for student involvement in the applied research process.

References

Alter, K. W., Barrows, A. K., Jennings, C. W., and Powell, J. D. (2000). 3-D perspective primary flight displays for aircraft. *Proceedings of the 44th Annual Meeting of the Human Factors and Ergonomics Society*, 3-29 – 3-32.

Arthur, J.J., III, Prinzel, L.J., III, Kramer, L.J., Parrish, R.V., and Bailey, R.E. (2004). Flight simulator evaluation of synthetic vision display concepts to prevent controlled flight into terrain (CFIT). Springfield, VA: NTIS, NASA/TP-2004-213008.

Beringer, D. B. (2000). Development of highway-in-the-sky displays for flight-path guidance: History, performance results, guidelines. *Proceedings of the 44th Annual Meeting of the Human Factors and Ergonomics Society*, 3-21 – 3-24.

Gershzohn, G., (2001). Unusual attitude recovery using the roll arrow. Society of Automotive Engineers, Inc., 641-5.

MIL-STD-1787C (revised 5 January 2001), Department of Defense interface standard: Aircraft display symbology. Wright-Patterson AFB, OH, ASC/ENOI.

Prinzel, L.J.,III, Hughes, M. F., Arthur, J.J., III, Kramer, L.J., Glaab, L.J., Bailey, R.E., Parrish, R.V., and Uenking, M.D. (2003). Synthetic vision CFIT experiments for GA and commercial aircraft: "A picture is worth a thousand lives". *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, 164-8.

Wickens, C. D., Haskell, Il, & Harte, K. (1989). Perspective format for a primary flight display. *Tech. Rpt. ARL-89-2/Boeing-89-1*. Savoy, IL: University of Illinois, Aviation Research Lab.

USE OF A TRAFFIC DISPLAY TO SUPPLEMENT VISUAL SEPARATION DURING VISUAL APPROACHES

Randall S. Bone, **John Helleberg,** and **David A. Domino**
MITRE
McLean, VA

At many busy airports, maximum efficiency and minimum delay occur when visual approaches are being conducted by pilots using visual separation from traffic. Pilot willingness to accept responsibility for visual separation also affords controllers maximum flexibility in traffic management under conditions of high traffic load. It may be possible to extend that efficiency to lower weather conditions if pilots are able to perform the same separation tasks by reference to a Cockpit Display of Traffic Information (CDTI) in lieu of visual contact out-the-window (OTW). This study is the third in a series of four designed to examine whether a CDTI can be used for this task. This particular study documents the first simulation to examine the concept during visual approaches. Eight commercial airline pilots flew visual approaches in a flight deck simulator, while maintaining a self-determined separation from the traffic, using two airspeed control methods: autothrottle and manual throttle. The objective and subjective results indicate that pilots are willing and able to perform this procedure (named CDTI Assisted Visual Separation (CAVS)) during visual approaches, using either the autothrottle or the higher workload method of manual speed control.

Background

A Cockpit Display of Traffic Information (CDTI) using Automatic Dependent Surveillance-Broadcast (ADS-B) has been identified as an integral element of the future Air Traffic Management (ATM) system (e.g., RTCA, 2002b). Following some of the early studies of CDTI (e.g., Connelly, 1977) and the first deployment of rudimentary traffic displays associated with the Traffic Alert and Collision Avoidance System (TCAS), standards for a more robust CDTI and an associated datalink have been developed (RTCA, 2003). Additionally, a set of operational applications for the use of CDTI also have been been identified (e.g., RTCA, 2002a). However, limited research has led to operational implementations of only a few near-term applications (e.g., Olmos, et al, 1998). The current study is directed at fielding one of the near-term applications using currently available avionics (i.e., Garmin AT2000, a CDTI system) and supporting the implementation by a customer who has installed the avionics in a portion of its fleet (i.e., United Parcel Service (UPS)). Initial development of the concept has been under the name CDTI Enhanced Flight Rules (CEFR). In the present study we adopt a more descriptive (and currently accepted) term, CDTI Assisted Visual Separation (CAVS).

Introduction

Visual separation can be used by Air Traffic Control (ATC) to separate aircraft in terminal areas by delegating responsibility to the flight crew who sees the other aircraft involved. If the flight crew accepts a clearance by ATC to maintain visual separation, it must: maintain constant visual surveillance, maneuver the aircraft to maintain in-trail separation, avoid wake, and notify ATC if visual contact with the other aircraft is lost.

When visual separation is to be used, a traffic advisory is issued by ATC to the flight crew. The flight crew then searches out-the-window (OTW) for the traffic and, when the traffic is visually acquired, reports it in sight. The search for aircraft in a dense traffic environment, during reduced visibility, or at night can be challenging (Stassen, 1998). The flight crew may have difficulty visually identifying aircraft and may even identify the wrong aircraft as the traffic of concern. These problems can be reflected in the number of traffic advisories that must be issued before the traffic is visually acquired, or the need for a controller to intervene to re-establish separation. After reporting the aircraft in sight, the flight crew is assigned responsibility for visual separation and a visual approach clearance can be issued. Thereafter, the flight crew is responsible for maintaining separation from the Traffic To Follow (TTF) to the runway, while ATC continues to provide separation from all other aircraft.

While maintaining visual separation, the flight crew must adjust spacing as necessary to maintain a safe arrival interval, and may have to detect and then respond to unexpected deceleration of the TTF, requiring them to adjust speed, reconfigure the aircraft, and in extreme cases perform a go-around (if the flight crew judges the separation to be unsafe). Detection of decreasing range to a visual target can be particularly difficult during clear nights when aircraft lighting blends with other ground lights. On occasion, the flight crew may lose sight of the preceding aircraft, requiring ATC intervention to establish another form of separation.

Experience with the TCAS traffic display, and formal studies, have shown that a display with traffic information is an effective enhancement to visual acquisition (Andrews, 1984). In fact, the concept of using a traffic display for enhanced visual acquisition is currently being practiced effectively in TCAS-equipped aircraft. During an operational evaluation of pilot use of CDTI, flight deck observers noted that when a CDTI was available to enhance airborne traffic awareness it was normally the first method used, followed by an ATC advisory or visual OTW sighting. This pattern of use occurred during day (with poor visibility) and night (with good visibility). In this flight test, approximately 75% of the traffic events involved use of the CDTI (Joseph, et. al., 2003).

The additional information available on the current generation CDTI may also allow the flight crew to make more accurate spacing judgments using features such as closure rate, speed and distance information, as well as a range ring with a spacing alert (see Figure 1). The absence of this information was noted during an assessment of the capability of the TCAS traffic display to support pilot managed separation (Hollister and Sorenson, 1990).

Finally, when losing sight of the aircraft, Imrich (1971) noted that the CDTI should assist in traffic awareness when transitioning in and out of clouds, at night, or during visual illusions. During an operational evaluation / flight test, flight crews reported that the CDTI helped in maintaining an awareness of the exact position of traffic when flying instrument approaches with visibility less than 5 miles and the TTF transitioned in and out of cloud layers (Battiste, Ashford, and Olmos, 2000).

If information on a CDTI can be used to supplement the visual separation task, visual approaches may continue to be used during conditions under which visual OTW contact cannot be maintained. Loss of visual contact would normally require that visual approaches be suspended with an associated loss of arrival capacity. The ability to continue visual approach operations under the proposed concept has been shown to be beneficial (FAA, 2003).

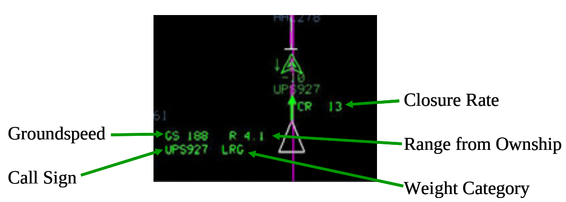


Figure 1. *Inset of CDTI showing an ADS-B selected target and the associated target information.*

Earlier studies in this series examined CAVS in the context of the instrument approach. Discussions among stakeholders and results of the previous simulations led to the conclusion that visual approaches will be the most likely initial implementation. This is due to several issues arising from use of a CDTI for separation in full instrument conditions (for a description of the instrument application, see Bone, Domino, Helleberg, and Oswald, 2003). In the present visual approach application, requirements for the conduct of the visual approach are unchanged except for pilot use of the CDTI to supplement visual separation. The flight crew will be required to correlate the aircraft seen OTW with the target on the CDTI prior to using the CDTI to maintain separation.

Method

The MITRE CAASD ATM simulation facility is an end-to-end, human- in-the-loop simulation consisting of a generic, fixed base, mid-fidelity transport cockpit with a visual display system, controller stations, pseudo-pilot capability, and the associated simulated radio communications. Confederates supported the simulation and provided simulated communication with the controller and the other aircraft inbound to the landing runway.

Subjects

Eight air carrier pilots (mean flight time = 8235 hours) were recruited for the study and were paid for their participation. All had glass cockpit experience and were familiar with the TCAS, which includes a rudimentary traffic display. All were currently flying turbojet aircraft. Each pilot acted as the “pilot flying” during the simulation. An air carrier qualified confederate acted as the “pilot not flying” and performed CAVS-specific duties such as interaction with the CDTI and providing verbal closure rate advisories.

CDTI

The CDTI was located in the throttle quadrant forward console area (the same location typically used in some weather radar installations). This location, out of the primary field of view, represented a lower cost retrofit location and a likely initial implementation. The CDTI display size had a 7 inch (17.8 cm) diagonal. This display size and location was shown to be acceptable in a previous simulation (see Bone, Helleberg, Domino, and Johnson, 2003).

The CDTI feature set was that required for the Enhanced Visual Approach procedure as defined in RTCA, 2003. Targets were displayed as chevrons. A specific target could be selected to display additional information. The available information included target ground speed (in one knot increments), range (in 0.1 nautical mile increments) from ownship, flight identification, and weight category (see Figure 1). Closure rate (in one knot increments) to TTF was automatically displayed when certain geometry constraints were met. Target range alerting was not provided. The traffic information was overlaid on the navigation display. Targets appearing on the CDTI were correlated with visible traffic in the external visual scene.

Procedure

The experiment used a single independent variable (method of speed control) in a "within subjects" design. Two speed control conditions were examined: manual control and autothrottle control. When using the autothrottles, speed commands were input through the Mode Control Panel (MCP). Without the autothrottle, pilots manually controlled airspeed using the throttle levers. The manual speed control condition was expected to produce higher workload. Method of airspeed control was counterbalanced across scenario events and all subjects experienced both methods.

At the beginning of the simulation, pilots completed a pre-simulation questionnaire, were briefed on the purpose of the study, and flew three practice approaches to become familiar with the simulation characteristics, CAVS procedure, CDTI features, and speed control.

After training, the data collection approaches began. Parallel visual approaches were in effect for runways 17 Right and 17 Left at the Louisville Standiford Airport (SDF), the main sorting hub of UPS. Subjects flew a visual approach to either runway, however the autopilot and approach coupler were used for flight guidance on the available instrument landing system. They were informed in advance whether the current approach would be flown with or without the autothrottle. Each trial began with ownship and TTF in clear conditions on either downwind or a dogleg to final on top of a haze layer. TTF weight category was varied, with large, Boeing 757 and heavy jet traffic simulated. Pilots would consider this information in selecting their desired minimum spacing. Final approach speed within TTF category was also varied, with speeds drawn from a distribution appropriate to that category. This reduced the ability of pilots to "learn" the final approach speed of a particular TTF category during the trials

The weather included a haze layer that began at 4000 feet Mean Sea Level (MSL) and continued down to ground level. This haze layer allowed for the visual acquisition of the TTF above the layer and assured the loss of the aircraft from the visual OTW scene during the final approach segment.

After each approach, pilots taxied clear of the landing runway and completed a workload form. Each approach spanned approximately ten minutes. After the simulation, pilots completed a questionnaire and participated in an informal debrief.

Data

One of the main purposes of this study was to assess pilot spacing behavior while using a CDTI to monitor spacing in a manner similar to that used while maintaining separation under a current visual separation clearance. The spacing and closure rate between TTF and ownship was collected at a rate of once per second after the simulation aircraft and the TTF were within, and remained within, the final approach corridor.

After each approach, pilots completed a Bedford Workload Rating Scale. The Bedford workload form is a modification of the Cooper-Harper measure of handling qualities of test aircraft. Pilots also completed a written questionnaire and debrief at the conclusion of the study.

Results

Objective Data

The final spacing data represented the in-trail separation as the TTF crossed the threshold, which is commonly used as one measure of throughput efficiency. To increase power and allow comparisons across TTF weight categories with a single ANOVA, the spacing data was converted into a relative measure of the distance between ownship and the radar separation minima (including wake turbulence requirements). It should be noted that required radar separation was used only as a reference point and is not required to be used by flight crews maintaining visual separation.

Due to various data collection issues, 15 trials were excluded from the objective data analysis. This yielded a total of 81 trials with usable objective data, which were used for the following analyses.

In order to assess the effect of speed control method on threshold spacing, a within-subjects ANOVA was conducted on the spacing data, the spacing between TTF and ownship at the threshold was not significantly affected by speed control F (1, 7) = 0.10, *ns*.

Pearson product-moment correlations between the initial spacing at the point at which pilots began the spacing task and the final spacing when TTF crossed the threshold were performed separately for each TTF weight category. The correlations between initial spacing and spacing at the threshold were significant for all three aircraft categories: large TTF r (41) = 0.86, p < .01, 757 TTF r (19) = 0.84, p < .01 and heavy TTF r (15) = 0.50, p < .05, indicating that spacing at the threshold increased as the initial spacing increased.

Figure 2 depicts the relationship between initial spacing and final spacing for TTF in the large weight category. The graphs for the 757 and heavy aircraft show similar trends. Due to the variability introduced by dynamic assignment of the TTF and cockpit pairs, and distribution of final approach speeds for each TTF, a number of the TTF aircraft reduced their final approach speeds to extremely slow levels. These extremely slow TTF aircraft are also depicted in Figure 2. Not surprisingly, the figure shows a trend indicating that when pilots were following the unexpectedly slow TTF the threshold spacing tended to decrease.

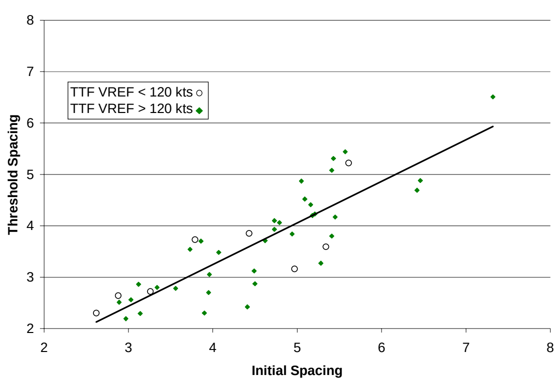


Figure 2. *Relationship between initial spacing and spacing when the TTF crossed the threshold when following large aircraft, with both nominal and slow final approach speeds*.

The effect of speed control method on minimum, mean, and maximum closure rate during each approach was also assessed using a within-subjects ANOVA.

The minimum closure rate was not significantly affected by speed control F (1, 7) = 0.55, *ns*. However, the maximum closure rate between TTF and ownship was significantly affected by speed control F (1, 7) = 5.55, p < .05. Similarly, the mean closure rate between TTF and ownship was also significantly affected by speed control F (1, 7) = 5.66, p < .05. Figure 3 shows that using autothrottle generally resulted in lower closure rates.

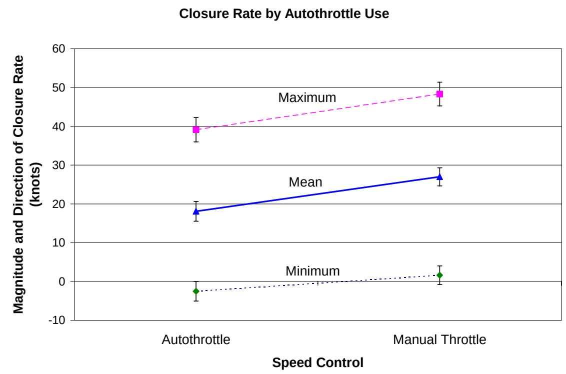


Figure 3. *Relationship between closure rate magnitude and speed control method.*

In order to examine the relationship between closure rate and the distance from TTF, the spacing data was converted into difference scores using the applicable radar separation minima for that weight category, (including wake turbulence considerations) as a reference point These derived values were used in the following analyses.

To assess the effect of speed control method on initial spacing, a within-subjects ANOVA was conducted on the initial spacing between ownship and TTF by speed control method. Initial spacing was not significantly affected by speed control F (1, 7) = 0.41, ns. Therefore, the observed closure rate differences between autothrottle speed control (M = 1.2, SD = 1.44) and manual speed control (M = 1.4, SD = 1.24) use were not due to differences in the initial spacing between TTF and ownship. Additional follow-up examinations of the data also did not indicate a reason for the closure rate differences for autothrottle and manual speed control.

Pearson product-moment correlations were performed between minimum, mean, and maximum closure rates and the derived spacing values at threshold. The correlation between minimum closure rate and distance from the spacing reference was significant r (79) = .35, p < .01. The correlation between maximum closure rate and distance from the spacing reference was marginally significant r (79) = .20, p = .08. The correlation between mean closure rate and distance from the spacing reference was significant r (79) = .37, p < .01. Figure 4 depicts the relationship between mean closure rate and distance from the spacing reference. It is clear from the figure that pilots utilized higher closure rates when the spacing between aircraft was greater and lower closure rates when spacing was reduced.

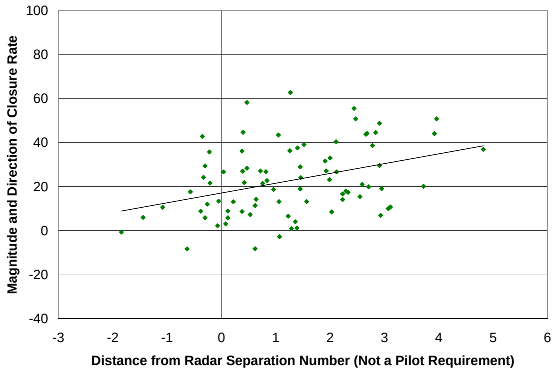


Figure 4. *Relationship between derived distance from spacing reference and mean closure rate across the entire approach.*

Subjective Data

All pilots reported that they would be willing to perform the separation task by sole reference to the CDTI under the conditions simulated in this study with either manual or autothrottle speed control. All pilots also reported that they were more confident when using the CDTI, versus OTW visual cues only, for establishing spacing. Pilots agreed that the necessary CDTI display features were available and those features (see Figure 1) were beneficial in performing CAVS.

Before starting data collection, pilots were asked to complete a baseline Bedford Workload Rating Scale estimating their workload during a typical visual approach while using visual separation. One of the pilots was unable to complete all 12 approaches due to a simulation malfunction and therefore, completed only 10 of the workload forms. This yielded a total of 94 with workload data.

The workload ratings provided by pilots at the end of each approach were subjected to a within-subjects ANOVA to assess the affect of speed control method on reported workload. The results revealed a significant main effect of speed control, F (1, 7) = 6.33, p < .05. Figure 5 indicates that regardless of speed control method, the overall workload ratings were similar to the baseline and remained relatively low (see Table 1).

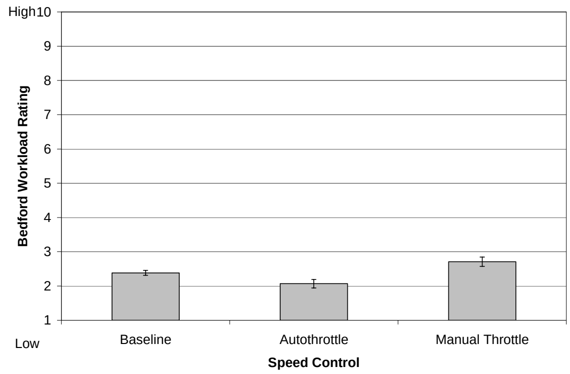


Figure 5. *Raw workload ratings for baseline visual approach and two speed control conditions while performing CAVS.*

Condition	Mean	Standard Error of the Mean	Standard Deviation	Bedford Workload Level
Baseline	2.4	0.07	0.70	Workload is low
Autothrottle	2.1	0.12	0.83	Workload is low
Manual throttle	2.7	0.14	0.94	Enough spare capacity for easy attention to additional tasks

Table 1. *Values of workload ratings in relation to two speed control conditions while performing CAVS.*

Conclusion

This evaluation of CAVS used the concept during visual approaches, with a CDTI located in the throttle quadrant forward console, and examined workload associated with different methods of speed control. It replicated the findings of the two previous studies in that pilots were able to adequately perform separation monitoring by reference to the CDTI with acceptable workload. (Bone, Domino, Helleberg, and Oswald, 2003; Bone, Helleberg, Domino, and Johnson, 2003).

Pilot responses indicated strong acceptance of the CDTI features. In fact, pilots reported being more confident with the use of the CDTI as compared to the OTW visual cues for establishing the appropriate spacing. Objective closure rate data indicated that pilots were able to use the information available on the CDTI to allow for higher closure rates when spacing between aircraft was greater and lower closure rates when spacing between aircraft was reduced.

When following all aircraft weight categories (large, 757, and heavy aircraft), final spacing between ownship and the TTF increased as initial spacing increased. These results indicate, as with the previous simulations (Bone, Domino, Helleberg, and Oswald, 2003; Bone, Helleberg, Domino, Johnson, 2003), that controllers will continue to have a key role in the successful implementation of CAVS, since their vectoring procedures will determine the initial spacing between aircraft on the approach. Tighter initial spacing or an instruction to maintain a certain speed or greater will permit pilots to "fine tune" their spacing intervals.

While the higher workload reported for manual speed control was statistically significant, it was not operationally significant since the workload rating on the Bedford Workload Rating Scale was still, "enough spare capacity for easy attention to additional tasks." Additionally, pilots reported being willing to perform CAVS while using either the autothrottle or manual throttle for speed control.

In the objective data, there was no effect of speed control method on final threshold spacing. However, higher closure rates were associated with manual speed control. These higher rates were found for the minimum, mean, and maximum closure rates (while only the mean and maximum were statistically significant). While there were differences between manual and autothrottle speed control for closure rate, final spacing was not affected, thereby indicating that the closure rate differences, while interesting, may not be operationally relevant. However, further investigation or simulation may be desirable.

CAVS is in the preliminary stages of development and evaluation. The final simulation in this series will again examine the visual approach application but during night conditions. The simulation will also examine the effects of automatic range alerts, failure conditions, and flight crew coordination procedures.

References

Andrews, J. W. (1984). *Air-to-Air Visual Acquisition Performance with TCAS II*, Project Report ATC-130, DOT/FAA/PM-84/17. Lexington, MA: MIT LL.

Battiste, V., Ashford, R. and Olmos, B. O. (2000). *Initial Evaluation of CDTI / ADS-B for Commercial Carriers: CAA's Ohio River Valley Operational Evaluation*. SAE Paper Number 2000-01-5520. Presented at the 5th World Aviation Congress and Exposition, San Diego, CA, October 10-12, 2000.

Bone, R. S., Domino, D. A., Helleberg, J., and Oswald, A. (2003). *Cockpit Display Based Visual Separation During an Instrument Approach: Pilot Performance and Acceptability. An Implementation of Cockpit Display of Traffic Information (CDTI) Enhanced Flight Rules (CEFR).* MTR 03W0000011. McLean, VA: MITRE.

Bone, R. S., Helleberg, J., Domino, D. A., and Johnson, N. (2003). *Effects of Traffic Display Size and Location During an Instrument Approach Spacing Task: Cockpit Display of Traffic Information (CDTI) Enhanced Flight Rules (CEFR) Simulation 2*. MTR 03W0000030. McLean, VA: MITRE.

Connelly, M. E. (1977). *Simulation Studies of Airborne Traffic Situation Display Applications*. ESL-R-751. Cambridge, MA: MIT.

FAA (2003). *Safe Flight 21 CDTI Enhanced Flight Rules (CEFR) Initial Benefit Analysis (Version 3)* May 2003, Washington, DC: DOT FAA.

Hollister, W. and Sorenson, J. (1990). *TCAS – Cockpit Display of Traffic Information (CDTI) Investigation Phase I Report: First Draft*. Sunnyvale, CA: Seagull Technology, Inc.

Imrich, T. (1971). *Concept Development and Evaluation of Airborne Traffic Displays*, MIT FTL Report R71-2. Cambride, MA: MIT.

Joseph, K. M., Domino, D. A., Battiste, V., Bone, R. S., and Olmos, B. O. (2003). *A Summary of Flightdeck Observer Data from SafeFlight 21 OpEval-2*. DOT/FAA/AM-03/2. Oklahoma City, OK: FAA CAMI.

Olmos, B. O., Mundra, A.D., Cieplak, J. J., Domino, D. A., and Stassen, H. P. (1998). Evaluation of Near-Term Applications for ADS-B/CDTI Implementations, *Proceedings of the 1998 World Aviation Conference*. Warrendale, PA: SAE.

RTCA (2002a). *Minimum Aviation System Performance Standards (MASPS) for Automatic Dependent Surveillance – Broadcast (ADS-B)*, RTCA/DO-242a. Washington, DC: RTCA.

RTCA (2002b). *National Airspace System: Concept of Operations and Vision for the Future of Aviation*. Washington, DC: RTCA.

RTCA (2003). *Minimum Aviation System Performance Standards (MASPS) For Aircraft Surveillance Applications (ASA)*, RTCA/DO-289. Washington, DC: RTCA.

Stassen, H. P. (1998). *Enhancing the Safety and Efficiency of Runway Approaches Using a Cockpit Display of Traffic Information*, WN98W0000005. McLean, VA: MITRE.

EID FOR A TERRAIN-AWARE SYNTHETIC VISION SYSTEM

Clark Borst, **Marieke Suijkerbuijk**, **Max Mulder**, **René (M. M.) van Paassen**
Delft University of Technology, Faculty of Aerospace Engineering, Control and Simulation Division
Delft, The Netherlands

Synthetic Vision Systems (SVS) are likely to become an integral part of the commercial flight deck in the future. The introduction of SVS is driven by the need to increase safety, most notably to reduce Controlled Flight Into Terrain (CFIT). Various avionics companies and research institutes have successfully developed SVS that have shown to increase the pilot's situational awareness regarding to attitude, position and clearance relative to the terrain. To further increase the pilot's terrain awareness, we believe that more meaningful information should be added to the synthetic view on the outside world. This can be accomplished by showing the pilot how the external constraints (terrain) relate to the internal aircraft constraints (e.g. climb performance). Based on that information, a pilot can see for himself what an obstacle actually means to him in terms of possibilities to fly over it, and if not, what his alternatives for action are. A guiding principle to develop a more meaningful interface is the paradigm of Ecological Interface Design (EID). This paper presents the preliminary results of an aviation work domain analysis conducted with respect to the manual control task of guiding aircraft through a terrain-challenged environment. This work will serve as the foundation for developing an ecological SVS interface with the objective to truly enhance the pilot's terrain awareness.

Introduction

The dominant factor in all aviation fatalities can be attributed to Controlled Flight Into Terrain (CFIT) accidents (Breen, 1997). Analysis conducted by the Flight Safety Foundation (FSF) showed that 90% of the CFIT accidents occurred in Instrument Meteorological Conditions (IMC) (FSF, 2002), which indicates that current aircraft safety and warning systems are inadequate in providing situational awareness (SA). In order to prevent these types of accidents, intuitive systems are needed that continuously inform the pilot about his/her spatial orientation in terms of terrain and flight path. Synthetic Vision Systems (SVS) are believed to provide these features, because the hypothesis is that when you show the picture, the pilot will get better awareness. However, recent research indicates that a SVS alone does not inform the flight crew accurately enough about their clearance relative to the terrain (Schiefele, Howland, Maris and Wipplinger, 2003). Therefore, a SVS is still backed by advanced terrain warning systems like the (Enhanced) Ground Proximity Warning System ((E)GPWS). These systems address this issue by providing warning messages and procedural tasks to be executed in order to avoid terrain collisions. They have proven to be of inestimable value in reducing the number of CFIT accidents (Figure 1). However, in combination with a SVS the warn-act strategy used by the (E)GPWS is not a very elegant solution. The warning messages and procedural tasks it supplies, force the flight crew to be reactive rather than proactive and this could decrease the SA. It would be better to have a SVS that graphically presents the meaning of the terrain towards conduction a safe flight. Hence, a better integration of the (E)GPWS functionality into the SVS is needed.

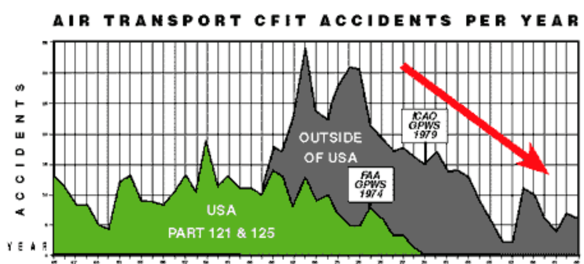


Figure 1 *The introduction of terrain warning systems such as the GPWS has reduced the number of CFIT accidents considerably.*

This paper investigates the possibility to use Ecological Interface Design (EID) to develop a SVS that adds more *meaning* to the computer-generated imagery of the outside world. This will be done by analyzing how the internal aircraft constraints, formed by its performance and maneuver limitations, relate to the external constraints formed by the terrain. Eventually, by visualizing the internal and external constraints on the SVS, the pilot will be much more aware of the margin within he can safely operate the aircraft.

The structure of this paper is as follows. First, the challenges that current SVS face are dealt with. Second, a definition for terrain-awareness is defined followed by the motivation for using the EID framework. Then, a test case in the vertical plane will be provided in order to analyze what is involved in flying over obstacles. Finally, the result of this analysis will be used to construct a preliminary AH of the manual control task when guiding an aircraft through a terrain-challenged environment.

Challenges of SVS

A SVS is basically a synthetic view of the surrounding world overlaid with essential aircraft status information (Figure 2). The main benefit of integrating all this information on a single interface is that pilots do not require diverting their visual attention away from external events and primary flight reference (Prinzel, Comstock, Glaab, Kramer and Arthur, 2004). Furthermore, it enables the flight crew to see the surrounding terrain even in low-visibility conditions. Therefore, SVS are believed to provide the adequate safety and SA enhancements needed to maneuver an aircraft through a terrain-challenged environment. By visualizing the terrain and obstacles ahead of the aircraft, the pilot can visually assess for himself whether or not an obstacle is a potential threat.

Figure 2 *SVS showing a perspective view on the surrounding terrain.*

Although a pilot can see the obstacles ahead of the aircraft, the SVS interface does not provide specific information what those obstacles actually mean to him. For example, the pilot sees on the SVS a mountain ridge at a certain distance ahead of the aircraft. What meaning has this mountain ridge to the pilot? Does it mean that the aircraft can fly over the ridge when it continues on the same course? If not, what kind of vertical maneuver will be required in order to fly over it safely? And at what moment in time should this maneuver be initiated? And if the aircraft will not be able fly over it due to its performance limitations, what kind of horizontal evasive maneuver will be required? Current SVS do not provide answers to these kinds of questions. They only show the pilots status and predictive information in terms of where they are and where they are going. Hence, the pilot himself is responsible for using his understanding of the aircraft's performance and its limitations in order to execute a feasible evasive maneuver. This task is further complicated by the relatively large Field Of View (FOV) adopted by many SVS, which makes it difficult to determine how close the aircraft is actually flying relative to the terrain and how fast the terrain is rising relative to the current altitude flown (Schiefele et al., 2003).

To give the pilot elementary meaning of the obstacles ahead of him, current SVS need to be equipped with Terrain Awareness Warning Systems (TAWS) or EGPWS. However, these warning systems were not designed to work specifically with a SVS interface. Therefore, the link between these systems and the SVS interface is not very elegant. Currently, when the EGPWS issues a caution, the caution is written as a message on the SVS interface (e.g. "Caution, Terrain" or "Terrain Ahead"). In case the EGPWS issues a warning, the warning message and what to do about it is also displayed on the SVS interface (e.g. "Terrain-Terrain, Pull Up-Pull Up"). It would be better to have a SVS that shows a graphical representation of the meaning of the terrain/obstacles ahead such that it will prevent the flight crew from ever coming in a hazardous situation where the EGPWS will be triggered. This requires the SVS to make the pilots aware of the aircraft's maneuver capabilities and limitations. Hence, the functionality of the EGPWS should be integrated into the SVS in order to increase the "terrain awareness" of the pilot.

Terrain Awareness

In general, keeping the SA of the flight crew at a high level is one of the most important jobs of the onboard aircraft systems. A definition for SA is '*the perception of the elements in the environment within a volume of space and time, the comprehension of their meaning, and the projection of their status in the near future*' (Endsley and Garland, 2000). Applying this definition to the pilot's awareness of the environment, he must be able to perceive the obstacles ahead, determine what those obstacles mean to him and make decisions based on that information. Current terrain warning systems automate the process of comprehending the meaning of those obstacles and making decisions how to act accordingly. The computer-generated decisions are then presented to the pilot in the form of tasks to be executed. Although procedural tasks can reduce the pilot's mental workload, it can also reduce his awareness about the situation at hand. Hence, in order to increase the terrain awareness of the pilot, the onboard systems should actually *support* the pilot's process of comprehending and decision making instead of automating and hiding them. Real terrain awareness will only be obtained by not only showing the obstacles, like a SVS currently does, but also by continuously showing the aircraft's performance and maneuver limitations such that a pilot can see for himself whether a situation is a threat to safety or efficiency, and can also see what possibilities and alternatives there are to escape from this. However, it can be expected that an EGPWS will still be needed as a warning system. But by adding meaningful information about the terrain and the aircraft's performance to the SVS interface, it can be imagined that an EGPWS caution/warning will hardly ever be triggered, and when it is triggered, the pilot fully understands why. A guiding principle to develop such an interface is the paradigm of Ecological Interface Design (EID).

Reasons for Using the EID Framework

EID is a theoretical framework for designing human computer interfaces for complex socio-technical systems. The term 'ecological' reflects the need for incorporating environmental constraints of the application domain into the design of an interface. It is important to mention that the framework describes more or less a number of guidelines to analyze the cognitive work domain rather than giving a specific recipe to determine what the interface should look like.

EID is originally developed by Rasmussen and Vicente (1992) to increase the safety in process control work domains like nuclear power plants. The EID framework has been applied successfully in the aviation domain for the design of a fuel and engine systems interface (Dinadis and Vicente, 1999) and an interface for the approach-to-landing (Amelink, Van Paassen, Mulder and Flach, 2003).

The goal of EID is to design interfaces that reveal the affordances of the work domain in such a way that they support each level of cognitive control. The property that makes EID so interesting is that it allows the operator to freely choose whatever means are available to solve a problem, or to apply any control strategy that satisfies the system goals based on the operator's preference and expertise. Furthermore, it assists the operator in constructing a mental model of the system. In contrast to interfaces based on procedural tasks, which only tell the operator what to do by giving *directions*, an EID interface provides a more convenient "*map*" of the system/situation so the operator can decide form himself what to do, how to do it and what his alternatives are. A well designed EID interface could even support the operator in coping with unanticipated events, which makes the interface more robust than interfaces or systems based on pre-programmed algorithms. Hence, this makes the EID framework a suitable candidate for designing a SVS interface or SVS overlays that will truly increase the pilot's terrain awareness.

EID for Supporting Terrain Awareness

System boundary

In order to successfully conduct a work domain analysis, a precise definition of the system's boundary is needed first. For this preliminary work the focus will be limited to the manual control task in the vertical plane of guiding an aircraft through a terrain-challenged environment. Therefore, the primary goal or "functional purpose" of the system (the aircraft) in the environment will be to safely operate it without colliding with terrain, or simply 'terrain avoidance'. In order to further analyze the work domain, the constraints that influence the system goals must be identified. These will primarily consist of external (terrain) and internal (aircraft) constraints. Most of the internal aircraft constraints have already been identified (Amelink et al, 2003). A brief summary of those results will be provided in the following text.

The Role of Energy in Flying

Pilots unconsciously act on the energy state of the aircraft in order to control it effectively. By experience, a pilot knows that he has enough room for safe maneuvering when he flies high and fast. From there, a pilot can safely exchange altitude to gain speed or the other way around (Langewiesche, 1944). They will especially avoid flying low and slow as this means that e.g. they do not have enough freedom to pull-up and gain altitude at the cost of speed in order to avoid obstacles or terrain. In essence, this mental model of maneuvering awareness is directly related to the awareness of the energy state of the aircraft. Hence, pilots like to have lots of total energy such that they have enough opportunity, as dictated by the law of conservation of energy, to exchange kinetic energy (speed) and potential energy (altitude) for maneuvering. This means that in the vertical plane the pilot essentially plays the role of energy manager of the aircraft.

Aircraft Manual Control Task

The aircraft manual control task with respect to energy has already been investigated. To manage the energy state of the aircraft, the pilot will generally apply two control strategies. In the first strategy, the throttle is used to control the vertical flight path (altitude) and the elevator to control speed. In the second strategy, the elevator is use to control the vertical flight path and the throttle to control speed. In terms of energy, the pilot actually controls with the throttle the total energy rate. The elevator is used to distribute the total energy between potential and kinetic energy. An abstract view of the manual control task (in the vertical plane) can be depicted as "the reservoir analogy" (Figure 3).

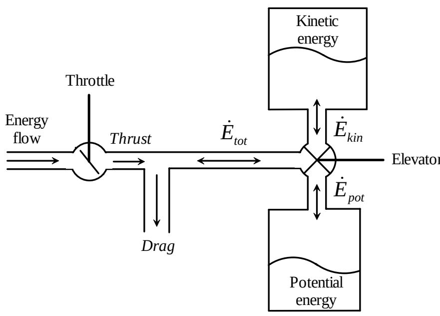


Figure 3 *The reservoir analogy, in which the throttle regulates the total energy flow and the elevator distributes the total energy flow between kinetic and potential energy (Amelink et al, 2003).*

Now that the aircraft manual control task is described in terms of energy, it remains to describe how this can help the pilot to maneuver over an obstacle. Clearly, the above analysis describes more or less the physics behind piloting itself, but it does not provide any information on how a pilot uses this to construct his mental model of the aircraft's maneuver capabilities to avoid terrain/obstacles. Therefore, in order to enhance the pilot's terrain awareness, he should continuously be confronted with the aircraft's performance and maneuver limitations based on its energy state.

The Role of Energy in Terrain Avoidance

With respect to terrain collision the position of the aircraft relative to the terrain is an important factor. Besides the position, also the aircraft's performance will play an important role. In the vertical plane it can be imagined that the energy state of an aircraft determines its climbing capabilities. Whether an aircraft is capable of safely passing an obstacle depends on the total amount of energy it possesses. If it is sufficient, enough kinetic energy can be exchanged by potential energy to be able to pass over the obstacle. This exchange is only limited by the minimum kinetic energy of the aircraft, referring to its minimum speed (stall). However, no aircraft is capable of exchanging its energy instantaneously. The exchange is bounded by the performance limitations of the aircraft and this also determines at what moment in time the pilot should initiate the evasive maneuver (Figure 4).

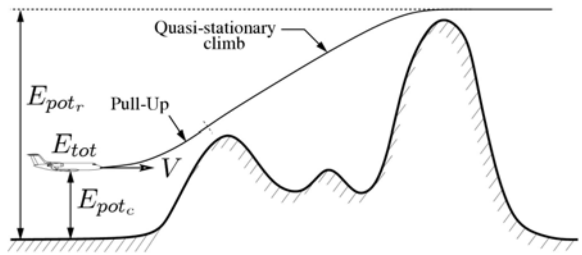


Figure 4 *How fast an aircraft is able to exchange its kinetic energy into the potential energy level that is required ($E_{pot,r}$) to safely pass over the terrain is limited by the pull-up maneuver and climb performance.*

Analysis showed that in the vertical plane three types of dynamic maneuver boundaries are important: the pull-up/pull-down maneuver, the optimal quasi-stationary climbing flight and the optimal gliding flight in case of total engine failure.

Performance Limitations

Pull-up/Pull-down Maneuver. As mentioned before, an aircraft will never be able to exchange energy instantaneously. When there is an excess (deficiency) of kinetic energy, a pull-up (pull-down) maneuver is used to initiate the exchange of energy. The pull-up or pull-down maneuver can be approximated by a circular maneuver (in the vertical plane). Analysis showed that when the vertical load factor of the aircraft will be limited to a certain value, the radius of the circle will increase with increasing speed. Hence, in high speed conditions, the pull-up maneuver will be important in avoiding terrain collision.

Optimal Climbing Flight. In general, there are three types of optimal climbing flights (Ruijgrok, 1996):

1. **The fastest climb** or least time to climb,
2. **The steepest climb** or minimum range during climb,
3. **The most economical climb,** where the smallest amount of fuel is consumed.

Here, the second type of climb is of highest concern since the functional purpose of the system is to increase safety and avoid terrain collisions at all costs. The steepest optimal climb will generally be executed by setting the thrust to climb-power and holding the indicated airspeed corresponding to this type of climb. This results in a maximum climb angle.

Optimal Gliding Flight. In general, there are two types of optimal gliding flights (Ruijgrok, 1996):

1. The gliding flight with the **longest duration** or flight at the minimum rate of descent,
2. The gliding flight resulting in the **maximum range** or flight at the minimum angle of descent.

Here, the second type of optimal gliding flight is of highest concern since it will not be interesting to know how long an aircraft is able to stay in the air. The optimal gliding flight will generally be executed by holding the indicated airspeed corresponding to this type of descent (typically, at which the drag is minimal).

The two optimal flights and the pull-up/pull-down will serve as the system's upper (climb) and lower (descent) performance boundaries (Figure 5). These boundaries can be used to detect a possible threat to safety and what the pilot can do to circumvent this threat and what his limitations are. For example, if a mountain rises steeper than the steepest climb angle reachable by the aircraft, the pilot is in trouble and should perform an evasive maneuver in the horizontal plane.

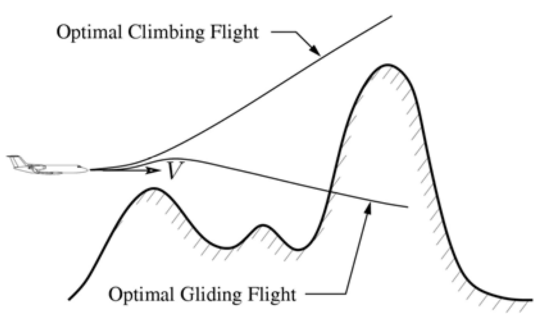


Figure 5 *The performance limitations can be used to detect a possible threat to safety. Here, the aircraft can still fly over the mountain ridge when initiating the optimal climb. However, in case of total engine failure an evasive maneuver in the horizontal plane will be required.*

A Preliminary Abstraction Hierarchy for Aircraft Terrain-Avoidance

In EID, the abstraction-decomposition space will serve as a representation of the work domain. The space consists of two dimensions, with along the top the *decomposition* (or *part-whole*) hierarchy and along the side the *abstraction* (or *means-ends*) hierarchy. In the decomposition space, each level represents a different granularity of the same work domain. Moving from left to right is equivalent to "zooming-in" because each successive level provides a more detailed representation of the work domain. The abstraction hierarchy ranges from, top to bottom, the most abstract level of purpose to the most concrete form of material. In general, higher levels in the AH represent the work domain in terms of its functional properties, whereas lower levels represent it in terms of its physical form.

The AH in this preliminary work will describe the work domain of aircraft terrain-avoidance in the vertical plane. The names of the levels are left the same as in Amelink's work. The content of the AH, for the analysis described in this paper, will be briefly discussed below and is summarized in Figure 6.

Functional Purpose

In general, the purpose of the system, i.e. the aircraft, in the environment is to fly to some destination and let it conduct a safe flight. Hence, the main goal is to reach the destination without colliding into terrain.

Abstract Function

This level describes the energy relations that govern the aircraft's movement in the vertical plane along with the energy of the terrain. In order to satisfy the goals of the level above, the potential energy constraint of the terrain and the aircraft's energy state are important.

Generalized Function

This level describes the aircraft maneuver functions and terrain shape function. The lift, weight, drag and thrust determine the constraints on the aircraft maneuver capabilities (pull-up/pull-down, optimal climb and optimal glide). The terrain's altitude profile determines the environmental constraint that the aircraft has to consider in order to satisfy the goals of the level above.

Physical Function

This level describes the physical implementation of the aircraft and terrain itself. They are the means that serve the ends of the level above. It includes the wings, control surfaces, power plant (engine) and the terrain's profile.

Physical Form

This level contains the geometry of the aircraft and the terrain's shape.

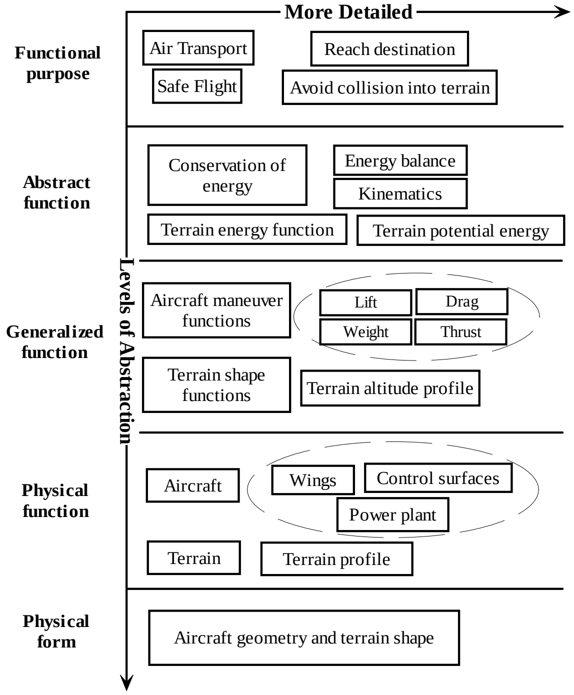


Figure 6 *A preliminary Abstraction Hierarchy (AH) for the aircraft manual control task in the vertical plane with respect to avoiding terrain collision.*

Conclusions

This paper can be considered to be work in progress. The preliminary AH has structured the problem of terrain collision avoidance in the vertical plane with respect to the external constraints (terrain) and internal aircraft constraints. The ultimate goal is to develop an ecological SVS interface that will assist the pilot in building a mental model of the aircraft maneuver capabilities in order to conduct a safe flight without colliding into terrain. The above analysis and AH reveals the dynamic aircraft maneuver limitations that has to become part of the interface. It is expected that the ecological SVS interface can be applied in a larger range of application domains than the EGPWS, because the analytical foundation of the interface's content contains more of the work domain.

The next step will be to evaluate a low-altitude terrain following task with a display concept based on the above analysis. Its purpose will be to determine to what extend the pilot is capable of avoiding terrain collisions with and without support by the interface.

References

Amelink, M.H.J., van Paassen, M.M., Mulder, M. and Flach, J.M. (2003). *Applying the Abstraction Hierarchy to the Aircraft Manual Control Task*. Proceedings of the 12th International Symposium on Aviation Psychology, April 14-17, Dayton (OH), USA.

Breen, B.C. (1997). *Controlled Flight Into Terrain and the Enhanced Ground Proximity Warning System*. The 16th Digital Avionics Systems Conference, October 26-30, pp. 3.1-1 – 3.1-7.

Dinadis, N. and Vicente, K.J. (1999). *Designing Functional Visualizations for Aircraft System Status Displays*. The International Journal of Aviation Psychology , Vol. 9, No. 3, pp. 241-269.

Endsley, M.R. and Garland, D.J. (2000). *Situation Awareness Analysis and Measurement*. SA Technologies. Lawrence Erlbaum Associates, Inc., Publishers. ISBN 0-8058-2134-1.

FSF (2002). *Reduced Visibility, Mountainous Terrain Cited in Gulfstream III CFIT at Aspen*. Flight Safety Foundation Accident Prevention, Vol. 59, No. 11.

Langewiesche, W. (1944). *Stick and Rudder: An Explanation of the Art of Flying*. McGraw-Hill Inc. (renewed 1072 by Wolfgang Langewiesche).

Prinzel, L.J. III, Comstock, J.R. Jr., Glaab, L.J., Kramer, L.J. and Arthur, J.J. (2004). *The Efficacy of Head-Down and Head-Up Synthetic Vision Display Concepts for Retro- and Forward-Fit of Commercial Aircraft*. The International Journal of Aviation Psychology , Vol. 14, No. 1, pp. 53-77.

Ruijgrok, G.J.J. (1996). *Elements of Airplane Performance*. Delft University Press, Delft, The Netherlands. ISBN 90-6275-608-5.

Schiefele, J., Howland, D., Maris, J. and Wipplinger, P. (2003). *Human Factors Flight Trial Analysis for 2D Situation Awareness and 3D Synthetic Vision Displays*. The 22nd Digital Avionics Systems Conference, October 12-16, Vol. 2, pp. 9.C.1 – 9.C.1-14.

Vicente, K.J. and Rasmussen, J. (1992). *Ecological Interface Design: Theoretical Foundations*. IEEE Transactions on Systems, Man and Cybernetics, Vol. 22, No. 4, pp. 589-606.

SPEECH ERRORS MANAGEMENT IN AIR TRAFFIC CONTROL COMMUNICATIONS: A DETAILED STUDY

Jean-Leon Bouraoui
IRIT/DIAMANT
Toulouse, France

Nadine Vigouroux
IRIT/DIAMANT
Toulouse, France

Many studies have reported on some human factors influencing the communication process, especially in aeronautical framework (see Davison (2003) for example). When spoken, communication comprises three different components: production, perception and understanding. The communication is often disturbed by one or many errors that affect one or several of these components. Consequently, one way to make air traffic control (ATC) communications more efficient and robust is to have as much knowledge as possible on these problems and their usual management. This paper presents the interests brought by corpus-based studies to Air Traffic Control (ATC) applications, especially interactions/communication between controllers and pilots. The corpus recorded represent dialogues during exercises where air-traffic controllers being formed interact/converse with people simulating pilots in practice. We propose error and strategies typology in accordance with the phraseology Then, we describe the principles and the specification adopted both for the recording and the annotation of corpus. Then, we report first results obtained from corpus analyses on errors and correction strategies of the air-traffic controller, and comment them in regards with ATC oriented applications.

Introduction

In the context of air controllers' activity, error handling is a very important thing, since it concerns the management of traffic and its security. The communication between air-traffic controllers and pilots must respect a phraseology (communication principles and rules).

We report how this handling is made during the air-controller formation. It consists to exploit a corpus of spoken dialogues that take place during air controllers' formation. We will show how this exploitation is made, via several levels of annotation (orthographic, semantic and dialogic) to study errors and corrections made during their formation. This goes through strategies of correction and self-correction. They are peculiar features of spontaneous speech, especially in stress and apprenticeship situation, as is the case with air controllers in formation. Indeed, because of the necessity of managing errors, each one has imperatively to be detected and corrected as soon as possible. We distinguish several categories of errors and different correction strategies.

In a first part, we will present the goal and the characteristics of the corpus, and the context in which it has been recorded. We will also comment/report the needs of a multi-layer annotation level for conducting natural language researches in the ATC domain. Then, we will present the annotation specification we chose for this work. Finally, we will give the results we obtained concerning errors and corrections and the categorizations it led us to.

Description of Corpus

Characteristics of controllers – pseudo-pilots communication

The formation of the Air Traffic Control (ATC) controllers includes theoretical teachings, but also consists of a lot of training sessions. These sessions are made of communication between air-traffic controllers being formed and "pseudo-pilots operators" (that is, people simulating pilots in practice).

The aim of the exercises is to train apprentice controller activities, and then evaluate them. It consists of managing several planes that are in a controlled area, for example by assigning them a given speed and/or position. Two languages were used: French and English (French being the majority). The exercise conditions were as near as possible from real environment: controllers worked with screen giving the radar position of virtual "planes"; the air traffic was simulated by several persons assuming the role of one or many pilots. Some background noises (overlapping conversations, sounds emitted by microphones, etc.) also occurred.

Figure 1 below is a formalization of the communication between a controller (C1) and a given pilot (pilot#1) until the controller addresses to another pilot (pilot#2).

The utterances produced by the controller, as well as the pilots' ones, must respect the phraseology. It describes, for example, the way the speaker must pronounce the planes call signs, or the order that the different components of a message have to follow. Two speakers can't speak at the same time, due to technical limitations: the audio channel is only assigned to one speaker. During the formation step, the phraseology is not always strictly respected even in real work conditions. But its general guidelines are kept. However, its learning and mastering was also aimed by exercises.

An instance of a simple order that an air controller can formulate to a pilot is: "Delta Tango Charlie climb level 9 0". We find, first, the call sign of the pilot's plane ("Delta Tango Charlie"), and then the order itself. In a regular grammar (Dourmap & Truillet, 2003), this utterance is composed by a call sign and the order. This last one is composed of a command, "climb", that plays the role of a predicate, whose argument is a value (for instance, "9 0" in our example). More complex utterances can also occur, composed of a sequence of simple orders. For a complete description of the French call signs and orders, see (Dourmap & Truillet, 2003).

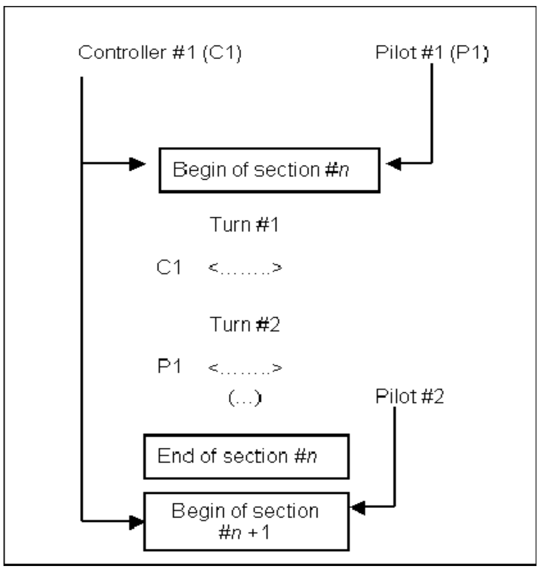


Figure 1: *Sections of sequences and turns*

Description of Corpus

The recordings were made on July 2001 at the ENAC (Ecole Nationale d'Aviation Civile; in English: National School of Civil Aviation) from Toulouse, in the framework of the VOICE[1] project. A DAT (Digital Audio Tape) was used. They were sampled at 16 kHz (16 bits). For recording reasons, the speech signal quality sometimes suffers from saturation or noises such as interferences. However, it stays intelligible. There were 16 speakers, and the total length of the corpus is 36 hours 50 minutes.

Transcription and Annotation Methodology

Multi-level annotation

According to the need, transcriptions and annotations of oral corpus can be opered at different levels:

1. Orthographic: putting what is said in writing, along with, possibly, the environment sounds. This level can also be augmented by labels of prosodic and extra-linguistic phenomena, such as pauses, hesitations, and so on;
2. Phonetical: transcribing what has been said in an I.P.A. (International Phonetic Alphabet). This level is useful to learn acoustic models for automatic speech recognition system and the various pronunciation of a word according (maternal language for instance).
3. Grammatical: assigning grammatical categorization to words of sentence. Some analysts also proceed to a lemmatization of words; that is to say, any inflected word is reduced to a canonical, basic form, called a lemma;
4. Semantical: this level can be processed according to different ways. For instance, one may seek to annotate words and/or sentence according to their meaning. On the other hand, the annotator can also focus his interest on the language acts expressed in sentences (in (Austin, 1962) sense). In the case of a corpus containing dialogs, such as our, it can also be the dialogs acts (Bunt, 1996) that are of interest. This kind of corpus can also be annotated according to a fourth level: dialogic one.
5. Dialogic: it concerns the structuring of the utterances produced by participants of dialogue. The annotation methodologies for this level are generally inspired from the works aiming to modeling dialogue and the combination of its components. One of the most famous is presented in (Roulet et al., 1985). To sum up, it consist in subdivide dialog in different

[1] Initially named VICTOR (Truillet & Vigouroux, 2001). VOICE goals are the study of spoken interaction utility and usability in the ATC area. To reach these aims needs: firstly to formalize under language models (like in (McTait et al. 2004) and (Dourmap & Truillet, 2003) for example) the phraseology used in real situation (Maugis, 1995); secondly to conceive a training environment where the pseudo-pilots will be replaced by spoken agents.

hierarchical levels. The main ones, from higher to lower, are: language act (the smallest unit), intervention (made by a given speaker, can be constituted of several language acts), and exchange (set of interventions related to a given topic).

As we have shown in this brief state of the art[2], there is a very large set of annotation methodologies. The choice is made according to the study aim of the corpus. We will show now in which way this study subject has led to the choice of a given way of transcribing and annotating.

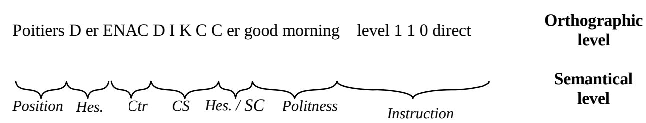
TranscriptionAannotation Methodology

We transcribed dialogues as well as annotated them according to some specifications ((Coullon & Gralia, 2000) and (Coullon et al., 2001)). The authors also made a distinction between the orthographic transcription and annotation. Annotation corresponds to an interpretation (at semantic, dialogic levels, etc.) of the orthographical string. These two tasks correspond respectively to the first, fourth and fifth levels described in the above multi-level annotation. Let's see more details.

Specifications are defined, firstly to determine elements that have to be transcribed. Secondly, to obtain homogeneity of transcriptions in case where several annotators processed the tasks. They consist essentially of rules to follow to transcribe technical ATC items such as call signs, speeds, etc. It also gives instructions to transcribe extra-linguistic events like hesitations, pauses, or accentuations. While transcribing the formation corpus, we believed that this specification wasn't sufficiently fine grained to mark out specific phenomena. Consequently, we contributed to the specifications by creating other classes of phenomena necessary to transcribe. We also refined existing one with sub-categories. Indeed, we considered the fact that the annotator could possibly not have access to the recordings, or not have time to refer to it for a given detail. Consequently, it is necessary to spot any phenomenon that could be interpreted as a marker for a language act, and accessible only via recordings hearing. For example, we introduced several tags corresponding to different pause lengths. This was based on the observation that, while a short pause could occur when one get his breath back, a longer one could spot something interesting in the speaker's behavior. For instance, he can have been disturbed by noticing he did an error, and seeking to fix it. We will come back on this example in the part devoted to correction study. In the same way, we noticed that frequently, the words produced when the speaker realize that he did an error are affected by a slight acceleration. Considering that this phenomenon could be considered like a marker of a correction, we decided to mark it with a special tag. It appears that, by doing this, we reach beyond the framework of "raw information" given by specifications. Indeed, this decision is based upon an interpretative act. However, we thought that if it wasn't done during the transcription, the annotator would miss some interesting phenomena.

We see here an illustration of the interconnection between the different levels of transcription/annotation we spoke about above. This lead us to the presentation of the transcription work.

As stated above, the aim was to give additional comments and labels to the transcribed elements. Thus, it would be possible to extract data according to a maximum number of criteria, and to carry out statistical researches (Coullon & Gralia, 2000, p.12). The information to give consists in two main categories. The first one corresponds to the identification of data related to the flights, like their coordinates, their ID, identity of speaker etc. The second aims to label the content of phrases, notably in terms of illocutionary function. This last category includes many fields. They marks for example opening and closing of dialog, politeness, or correction. In the second part of this article, we describe the study made on this last illocutionary act.



Caption : *Ctr*: Center ; *CS*: Call sign ; *Hes.*: Hesitation; *SC*: Self Correction

Figure 2: *Annotation of a simple order at two levels*

Work tool

The tool we used for transcription is Transcriber. It is a software developed at the DGA (Délégation Générale pour l'Armement: in English; General Delegation for Armament) to permit the transcription of broadcast (Barras et al., 2000). It offers advanced functions of transcription and annotation. It also allows to align transcription on signal. Furthermore, Transcriber gives opportunity to save transcription under several electronic formats, among which XML[3]. This last format is conceived to be easily portable and handled.

[2] For a more detailed overview, interested readers can confer to (Truillet & Vigouroux, 2001). Many works have been made on corpus; one of the nearest from our is (McTait et al., 2004).

[3] eXtensible Markup Langage.

Its usage is especially appropriate since this format has precisely been chosen to structuring the data obtained after the transcription of our corpus. Moreover, a DTD corresponding to the specifications was elaborated (Coulon et al., 2001). This DTD was completed by our added specifications.

These possibilities allow to simplify statistical enquires, such as counting the number of occurrences of the various strategies.

Exploitation of Transcribed Corpus: Application to Errors and Corrections Study

In a previous study (Bouraoui et al., 2003), we presented a complete study on this topic[4]. It is not the main subject of the present article. Consequently, we will only give the most outstanding results. Indeed, our aim is to illustrate the interest of this kind of work for the study of interaction between controllers and pilots. First, we present the categorizations we made, and conclude by giving the main results and comments.

Errors typology

After several viewings of the corpus, we noted that, whatever the error is, it's not the whole utterance (simple or complex, as definedabove) that is wrong, but only a part of it, or the way it is constructed. Due to this observation, we defined the following classes of errors:

- On an attribute: we mean by "attribute" an alphanumeric data that can be considered as an argument of a command. It can be for example a plane call sign ("Britair 452"), a position ("9 0"), a town ("Paris"), etc;
- On a command: a term (most often corresponding to an order, such as "climb", "request", etc.) is substituted to another;
- On utterance structure: a word or a group of words is not at its correct position in the utterance. For example "Air France 41 82 good morning climb level identify climb level 140": here, the speaker realized that he began to give the order "climb level 140" before the order "identify". Consequently, he corrects himself. The phraseology imposes the respect of the structure;
- On the language used: the speaker notice (or is being noticed) that he does not speak in the correct language (French instead of English or vice versa). For example, in the following dialog, the pseudo-pilot reminds to the controller that he must talk to him in English: Controller: "November 9 O O euh Fox Roméo contact ENAC 123 décimale 8" – Pseudo-Pilot: "in English please". This category is totally dependant of the ATC domain. Indeed, it is due to the fact that the controller has to speak one language according to the pilot he addresses to.

When an error is noticed, whether it is by the speaker or his interlocutor, it gives rise to various strategies of correction and self-correction, which we describe below.

Correction and Self-correction Strategies

We'll make a distinction between three main strategies of correction: self-correction of an element of the utterance being produced (either attribute or order), self-correction of a previous utterance, or correction coming from the interlocutor. The distinctive features of these categories are based on the person who does the correction (speaker or interlocutor) and the moment when it occurs. Indeed, we think that these different kinds of corrections can occur in distinct ways, and consequently be characterized by specific markers. Some studies on others oral corpora (notably (O'Shaughnessy, 1992), (Nakatani & Hirschberg, 1994), (Bousquet, 2002)) also revealed the existence of a phenomenon called "false-start" It occurs when the speaker begins a word, and stops producing it before the end. We considered it like an other category of self-correction.

Here follow examples of each of these categories, taken from our corpus (we set the element being corrected in italics):

- Self-correction: "KLM er 2 1 5 climb level 1 9 0 contact ENAC 120 contact ENAC er *1 2 6 decimal 8 5*.". The controller asks to pilot to go to level 190, and to contact ENAC on frequency 126.85. He makes a correction on the frequency to use. A particular kind of self-correction is false-start. For example: "Fox Golf Hotel Mike November ENAC good morning (…) speed minim er 200 Knots *minimum*.". The speaker begins to utter the word "minimum", and stops himself before ending it for he noticed that he did not give the speed;
- Correction of a previous utterance: here is a short dialog between a controller and a pseudo-pilot: Controller: "er Fox Kilo Charlie maintain level 1 7 0."-Pseudo-Pilot:"to level 1 7 0 Kilo Charlie." - Controller: "er Fox Kilo Charlie correction maintain level *1 9 0*." The controller first gives a position to which the pseudo-pilot must go. The pseudo-pilot confirms, but afterward, the controller

[4] Based on the two thirds of our corpus that were processed at that time. The present study is based on the whole corpus.

corrects his previous order, that was giving wrong coordinates;

- Correction from the interlocutor: here again, a dialog between a controller and a pseudo-pilot: Controller: "euh TAT 289 Mike Lima (…) join Poitiers" - Pseudo-Pilot: "Lacan Amboise Poitiers it's TAT Mike *India*.". In this example, the controller made a mistake on a part of the call sign of his interlocutor. Consequently, this one corrects him.

Markers

This part will be subdivided in two: we will first make general remarks about the different markers picked out, and then focus on the case of lexical ones, which present some interesting features.

General remarks. Two questions rise when one speaks about makers of a given phenomenon: what is the length of the scope around the phenomenon where something can be considered as marker, and which are the kinds of markers searched. Here are the principles we observed after viewing the corpus:

- We fixed the scope to 3 words before and after the correction phenomenon itself; this value results from empirical observations, as well as from the fact that some three "words" sequences form in fact the call signs; for more details on that point, see (Dourmap & Truillet, 2003);
- Three classes of markers were used: lexical, accentual and finally spontaneous speech phenomena. The two last ones results from the oral nature of the corpus: we employ the term "accentual" to designate the emphasis put on a word by the means of a variation of prosodic features (intensity for example). Thus, when a speaker corrects a wrong element within a call sign, it arrives that the element being corrected is pronounced with a particular accent. Let's take for example "Lacan Amboise Poitiers it's the TAT Mike *India*" (previously mentioned). The element in italics, that corrects a wrong value previously given, has been accentuated by the speaker, The class of "spontaneous speech phenomena" puts together various phenomena such as hesitations, repetitions (contrary to (Shin et al., 2002), we didn't put them in a specific category), or pauses. We call pause a non-speech period during more than half a second. We formulated the hypothesis that a silence during such a length is revealing of an enunciation problem such as the thought time necessary to find the correct word to say.

Lexical Markers. Among the lexical markers, we made the following classification, from what we observed:

- Deictic: word referencing to other word, such as "it's" (or "c'est" in French). The most frequent configuration is the following: "it's *CS*" (where *CS* is a call sign; for instance: "it's Alpha Mike Lima 753"). One should note that this usage of deictics are also quite frequently used in other contexts, especially by pilots to introduce themselves;
- Excuse: for example, "sorry", "excuse me", etc.;
- Negation: any words used in order to negate something, the most common one being "no";
- Correction: the word "correction". Its usage is explicitly asked by the phraseology for marking the correction of an utterance. It is also mentioned that the correction must be followed by the element corrected. Due to its status in phraseology, we put it in specific category.
-

Results and Comments

We'll display our statistics according to the classification presented above: firstly errors, then correction and self-correction strategies, to conclude with their markers.

Errors

On table 1, the reader will find the number of occurrences and the percentage (calculated in comparison with the total number of errors) of each category.

	Number	Percentage
Attribute	132	51,36%
Command	93	36,19%
Utterance structure	11	4,28%
Language	21	8,17%

Table 1: *Number and percentage of errors categories*

There's the same number of noticed errors that of corrections. (see also table 2). This is normal: any error has to be corrected at a moment or another, the sooner being the best. Most of the errors concern what we called "attribute", along with "commands". It is not surprising. Nearly all utterances contain at least one reference to a call sign, a speed, etc. The same reasoning can be applied to "commands". However, there is 1.5 times less errors committed on "commands" than on "attributes". This can be explained by the fact that "attributes", especially call signs and positions, are quite complex sequences of numbers and letters. Furthermore, they are only used in ATC context. Consequently, they certainly require handling an important cognitive load, thus leading to more errors. The cognitive load is all the more high since the apprentice controllers are in formation. This also explains the lesser number of errors of command utterances (nearly two times less occurrences than for “attributes”) and of structure (more than six times less occurrences than for “attributes”).

Corrections and Self-corrections

In table 3, we display the number of occurrences of the different kinds of correction found in the corpus. We also give their percentage in comparison with the number of speech turns. This last result must be tempered. Indeed, there are sometimes several corrections occurrences for one speech turn. In spite of this, it gives a good idea of the global proportion of this phenomenon through the corpus.

	Number	Percentage
Self-Correction	232	90,27%
Self-Correction of a previous utterance	16	6,23%
Correction by interlocutor	9	3,50%

Table 2: *Number and percentage of corrections strategies*

It appears that the most frequent kind of correction is the first one: the speaker corrects himself, during his current utterance. We now compare this result with those obtained a corpus of train reservations (Kurdi, 2003). The author count 241 self-corrections, on a total of 5300 speech turns[5]. In proportion to our corpus size, that makes a lot more self-correction occurrences in this corpus than in our. Lets examine this from a psycholinguistic point of view. It is admitted by most of authors (notably (Reason, 1990, p. 156 sq.) or (Levelt, 1999)) that, in the end of the speech production process, the locutor proceed to a “control” of what he actually said, in comparison to what he intended to say. In controllers’ production, this “control” is obviously more efficient that for people who does a “daily” task. Here again, we think that the responsibilities that the controllers does have enhance their attention to what they said.

Conclusions and Perspectives

We have studied a corpus of spontaneous speech dialogues, consisting of interactions between air controllers in formation and “pseudo-pilots”.

We shown, first, that the transcription and annotation of this kind of corpus is a very complex task. Its realization depends on the exploitation planned. Then, we detailed the methodology we applied. We chose it in order to constitute a structured data base in XML format.

In a second time, we sought to present the interest of corpus based works to study different sides of the ATC interactions. As a concrete illustration, we gave the main results of a previous study on errors and corrections in our corpus. It appears that the most frequent kinds of errors concerns what we called “attribute”, such as callsigns. We linked this to the fact that memorizing values need an important cognitive load, especially for novice controllers.

More generally, we saw that phraseology plays an important role for some of the errors that occur. For example, it is the case when the cause is a deviation regarding to the organization of the utterance.

In order to further explore this analysis, we plan to follow the two main ways we presented in this article. On one hand, setting up an enhanced methodology of transcription and annotation, sufficiently robust to be implemented into an automatic or semi-automated system, for example thanks to CACAO system (Bousquet, 2002). On the other hand, continuing our study on management of errors and their corrections. We could do this by leading cognitive studies on the notion of “attribute” and its cognitive load. A comparison between the apprenticeship dialogs we have with real ATC situations ones could also be done. This would benefit to one of the goals of VOICE projects, i.e. the implementation of communicating agents that would help pseudo-pilots and more generally to all researches concerning speech in ATC.

Acknowledgments

This study is funded by the CENA. We thank Philippe Truillet for the records of the exercises, and the information he gave us about it.

[5] (Kurdi, 2003, p. 74-75).

References

Arrêté du 27 juin 2000 relatif aux procédures de radiotéléphonie à l'usage de la circulation aérienne générale, J.O n° 171 du 26 juillet 2000, p. 11501.

Austin J. L. (1962) *How to do Things with Words*, Oxford University Press

C. Barras, E. Geoffrois, Z. Wu, and M. Liberman (2000). Transcriber: development and use of a tool for assisting speech corpora production, *Speech Communication special issue on Speech Annotation and Corpus Tools, Vol 33*, No 1-2.

Bousquet, C. (2002) *Compréhension robuste de la parole spontanée dans le dialogue oral homme-machine – Décodage conceptuel stochastique*, PhD Thesis, University of Paul Sabatier, Toulouse, France.

Bousquet-Vernhettes, C., Privat, R., Vigouroux, N. (2003) Error handling in spoken dialogue systems: toward corrective dialogue, *ISCA workshop on Error handling in Spoken Dialogue Systems*, Château-d'Oex-Vaud, Suisse, p. 41-45, 28-31 august 2003, ISCA, Stockholm.

Bouraoui J.-L., Vigouroux N., Bothorel G. (2003) Transcription and annotation of an apprenticeship corpus: application to the correction and self-correction strategies, *ISCA workshop on Error Handling in Spoken Dialogue Systems*, p. 35-40, Chateau-d'Oex-Vaud, Suisse, 28-31 august 2003. ISCA, Stockholm.

Bunt, H. (1996) Dynamic interpretation and dialogue theory, *The structure of multimodal dialogue*. Taylor, M. M., Bouwhuis, D. G. and Neels F., editors, John Benjamin, Amsterdam.

Coullon I., Graglia L. (2000) Spécifications de la base de données pour l'analyse des communications VHF en route, *CENA internal report*.

Coullon I., Graglia L., Kahn J., Pavet D. (2001) Définition détaillée du document type (DTD) pour le codage sous XML des communications VHF en route – VOCALISE Trafic CRNA / France 2000, *CENA internal report*.

Davison J. (2003) When Language Becomes a Barrier Instead of a Bridge: Communication Failures Between Pilots and Air Traffic Controllers, *12th International Symposium on Aviation Psychology*, Dayton, April 14-17 2003.

McTait K., Maynard H, Devillers L., Rosset S., Paroubek P., Mostefa D., Choukri K., et al.., (2004) Constitution d'un corpus de dialogue oral pour l'évaluation automatique de la compréhension hors -et en- contexte du dialogue, *JEP 2004*, Fès, Maroc, p. 357-360, 19-22 avril 2004.

Dourmap, L, Truillet, T. (2003) Interaction vocale dans le contrôle aérien : la comparaison de deux grammaires contextuelles pour la reconnaissance des indicatifs de vol, *CENA internal report*, NR03-669.

Kurdi M.- Z. (2003) *Contribution à l'analyse du langage oral spontané*, Ph.D. Thesis, Université J. Fourier, Grenoble, France.

Levelt w. J. M., roelofs a. Meyer a. S., A theory of lexical access in speech production, *Behavioral And Brain Sciences*, *22*, p. 1–75, 1999.

Maugis L., (1995) VOCALISES Speech Interface: Preliminary Studies, version 1.0, *CENA note*, N95051.

Nakatani C. H., Hirschberg J. (1994), A corpus–based study of repair cues in spontaneous speech, *Journal of Acoustical Society of America*, *953*, p. 1603-1661.

O'Shaughnessy D., (1992), Analysis of false starts in spontaneous speech, *Proceedings of the International Conference on Spoken Language Processing*, p. 931-934, Banff.

Reason, J., Human error. Eds: Cambridge University Press, 1990.

Roulet E., Auchlin A., Moeschler J., Schelling M., Rubattel C. (1985) *L'articulation du discours en français contemporain*, Berne, Lang (Collection Sciences pour la communication).

Shin J., Narayanan S., Gerber L., Kazemzadeh A., Byrd D. (2002) Analysis of user behaviour under error conditions in spoken dialogs, *Proceedings of the Seventh International Conference on Spoken Language Processing*, Colorado, September, 2002.

Truillet P., Vigouroux N. (2001) Interaction vocale dans l'ATC : le projet VICTOR, *IHM-HCI'2001*, p. 179-182, Lille, 10-14 septembre 2001.

Veronis J., (2000), Ingénierie des langues, Chapter "Annotation automatique de corpus : panorama et état de la technique", pages 111-129, Hermes, Paris.

LOW-AIRSPEED PROTECTION FOR SMALL TO MEDIUM-SIZED COMMERCIAL AIRPLANES: AN IMPORTANT SAFETY GAP

William J. Bramble, Jr.
National Transportation Safety Board, Office of Aviation Safety[1]
Washington, DC

Loren S. Groff
Charles M. Pereira
National Transportation Safety Board, Office of Research and Engineering
Washington, DC

In November 2003, the National Transportation Safety Board recommended that the Federal Aviation Administration (FAA) convene a panel of aircraft design, operations, and human factors specialists to examine the feasibility of requiring the installation of low airspeed alerting devices on airplanes operating commercially under 14 C.F.R. Parts 121 and 135. The Board further recommended that if the panel determined such a requirement to be feasible, the FAA should establish requirements for low-airspeed alert systems. This paper discusses the reasoning behind these recommendations, explores relevant accident history from the Safety Board's investigative records, and discusses shortcomings of an approach to cockpit design that relies on flight crew monitoring and artificial stall warnings for avoidance of low airspeed related accidents. Potential benefits and concerns associated with the installation of a new kind of low airspeed alerting device are also addressed.

Introduction

On October 25, 2002, a Raytheon King Air A100 on a non-scheduled Part 135 flight crashed 1.8 miles short of the runway threshold during a VOR approach to the Eveleth-Virginia Municipal Airport, Eveleth, Minnesota. Radar and weather data indicated that the flight crew experienced difficulty intercepting the approach course and performed a steep, fast approach, which probably required them to reduce engine power to very low levels. As the crew descended, their airspeed slowly and steadily decreased until it fell below recommended approach speed. Airspeed continued to decrease at a rate of approximately 1 knot per second for the last 48 seconds of flight. As the airplane reached the minimum descent altitude in the landing configuration, with its airspeed having decreased to near the calculated stall speed, the airplane suddenly rolled left, descended steeply, and impacted terrain. All occupants were killed, including the late U.S. Senator Paul Wellstone. The Safety Board found that icing was not a factor, and determined that the probable cause of this accident was "the flight crew's failure to maintain adequate airspeed, which led to an aerodynamic stall from which they did not recover" (National Transportation Safety Board, 2003).

In its final report on this accident, adopted on November 18, 2003, the Safety Board urged the FAA to convene a panel of aircraft design, aviation operations, and aviation human factors specialists, including representatives from the National Air and Space Administration to determine whether a requirement for the installation of low airspeed alert systems in airplanes engaged in commercial operations under 14 Code of Federal Regulations Parts 121 and 135 would be feasible (NTSB Recommendation No. A-03-53). The Board further recommended that if the panel determined such a requirement to be feasible, the FAA should establish requirements for low-airspeed alert systems (NTSB Recommendation No. A-03-54). This paper discusses the reasoning behind the Safety Board's recommendations, explores relevant accident history from the Board's investigative records, and discusses shortcomings of the current cockpit design philosophy relying on flight crew monitoring and artificial stall warnings to avoid low airspeed related accidents. Potential benefits and concerns associated with the installation of a new kind of low airspeed alerting device are also addressed.

Background

Airspeed is a basic measure of airplane performance monitored by flight crews. Angle of attack is the angle between the chord line of an airplane's wings and the oncoming relative wind. All other things held constant, when airspeed decreases, angle of

[1] The opinions expressed in this paper are those of the authors and do not necessarily reflect the official views and opinions of the U.S. National Transportation Safety Board.

attack must be increased to maintain lift. However, if angle of attack is increased too much, critical angle of attack can be exceeded, smooth airflow over the wing will be disrupted, and an aerodynamic stall results. A stall can occur at any airspeed, attitude, or power setting, however, if airspeed is allowed to decrease too much, a stall will reliably be produced.

Practicing aerodynamic stalls, and their recovery, is a routine part of pilot training. However, inadvertent stalls can be dangerous. This is especially true during the takeoff, climb, approach, and landing phases of flight. Inadvertent stalls are more likely during these phases because operating airspeeds are lower and stall speed margins are reduced. In addition, lower altitudes make stall recovery less certain. Flight crew airspeed monitoring is the first line of defense against inadvertent stalls. To guard against them, flight crews are trained to monitor airspeed instruments and to maintain target airspeeds.

Stall warnings provide a second line of defense against inadvertent stalls, serving as a backup to crew monitoring. Federal airworthiness standards (14 C.F.R. Parts 23 and 25) require the presence of a clear and distinctive warning capable of alerting the crew of an impending stall. This warning cannot require the crew's visual attention inside the cockpit, and must begin 5 or more knots above stalling speed for normal and commuter category airplanes. For transport category airplanes, it must begin at least 5 knots or 5 percent above stalling speed (whichever value is greater).[2] If the aerodynamic qualities of an aircraft (e.g., buffeting) do not provide a clear and distinctive warning meeting these requirements, an artificial stall warning must be installed. Flight crews are trained to begin stall recovery procedures if a stall warning occurs during normal flight operations.

The widespread introduction of swept-wing jet aircraft in commercial aviation in the 1960s brought an increased emphasis on stall avoidance, because stall recovery in such aircraft can be difficult or impossible (Federal Aviation Administration, 2004). Stick "pushers" installed on such airplanes were designed to lower the nose before critical angle of attack was exceeded, and artificial stall warning systems were required to be calibrated to activate at least five knots above stick pusher activation thresholds. Additional stall protection measures were developed in the late 20th century as manufacturers of fly-by-wire transport category airplanes with integrated autoflight systems developed flight envelope protection systems to prevent airplanes from exceeding high or low airspeed limitations. Full authority envelope protection systems, such as those installed on the Airbus A320, were made capable of increasing engine power and even modulating the effects of pilot control inputs to prevent exceedence of the critical angle of attack (Vakil, 2000).

The Safety Gap

Despite advances in the state of the art in stall avoidance and protection systems, many small to medium-sized commercial turboprop and turbine engine airplanes in use today still rely solely on flight crew monitoring and artificial stall warnings to avoid low airspeed-related accidents. This approach is problematic for two reasons. First, flying involves the time-sharing of multiple concurrent tasks, many of which require flight crews to monitor multiple displays. These tasks cannot always be performed simultaneously. For this reason, successful flying depends on effective prioritization and visual scanning strategies (Wickens, 2003). The process by which flight crews allocate their attentional resources among concurrent flying tasks has been called "cockpit task management" (Funk, 1991). Crews must ensure that important flying tasks, such as airspeed monitoring, receive adequate attention at appropriate times and are not pre-empted by lower priority tasks. Research has shown that pilots are generally good at doing this. However, a variety of evidence indicates that suboptimal cockpit task management does sometimes occur and can have a negative impact on safety (Wickens, 2003). Of interest to the topic at hand, the authors of one early study of flight crew performance in a full mission flight simulation cited violations of airspeed limitations (both high and low) as one of the most common types of flying errors made by three-pilot airline crews (Ruffel Smith, 1979).

A second problem with relying on pilot monitoring and stall warnings for stall avoidance has to do with characteristics of the stall warning itself. In theory, stall warnings are designed so that flight crews can prevent a stall by responding quickly to the occurrence of a stall warning. Current airworthiness requirements for transport category airplanes even state that it must be possible for a test pilot to prevent a stall during powered 1.5 G banked turns when stall recovery is delayed for at least one second after the onset of a stall warning. However, certain combinations of power changes and abrupt maneuvering (such as a level-off at MDA with or without structural icing) could reduce this margin of

[2] This requirement is reduced to 3 knots or 3 percent above stall speed when flying straight and level at idle power.

warning. Moreover, stall warnings can be unreliable because of ice accumulation, which raises stall speed and can degrade warning margins to the point where little or no warning is provided. This phenomenon was noted during the investigation of a 1997 accident near Monroe, Michigan that caused the deaths of 29 people, and led the Safety Board to recommend that the FAA apply more stringent certification requirements to airplanes certified for operation in icing conditions (National Transportation Safety Board, 1998).

Low Airspeed / Stall Events

In light of known human monitoring weaknesses and the potential inadequacy of artificial stall warnings, it should come as no surprise that the Safety Board has investigated numerous accidents and incidents involving flight crew failure to monitor and maintain airspeed. In some cases, loss of airspeed / stall events have been preceded by aggravating factors such as aircraft equipment or system failures that made airspeed monitoring and maintenance more difficult. Weather has also been an important contributing factor for low-airspeed related events. Aerodynamic stalls have occurred following encounters with wind, turbulence, and convective phenomena such as wind shear or microburst. However, structural icing may be the most common contributing factor.

Events involving flight crew failures to monitor airspeed can occur during any phase of flight, as the following example attests. On June 4, 2002, a Spirit Airlines McDonnell Douglas MD-82 on a scheduled Part 121 flight from Denver, Colorado to Ft. Lauderdale, Florida experienced an aerodynamic stall while cruising at 33,000 feet on autopilot. Fifteen minutes into the cruise phase of flight, the crew felt a sudden vibration, heard the stick shaker and stall warning activate, noticed that their airspeed was low and their engines were operating at a very low power setting. They also noticed that one engine's temperature was too high. The captain took manual control of the airplane and shut down the hot engine. Shortly thereafter, power rolled back on the good engine as well. The flight crew managed to restore power to both engines at 17,000 feet, and made a precautionary landing. The Safety Board found that the airplane's engine inlet probes had become blocked by ice crystals resulting in a false engine pressure ratio indication and subsequent retarding of the throttles by the auto throttle system. The Board attributed the probable cause of this incident to the flight crew's failure to verify the engine instrument indications and power plant controls while on autopilot with the auto throttles engaged, and their failure to recognize the drop in airspeed which led to an aerodynamic stall associated with the reduction in engine power (Safety Board No. CHI02IA151).

The authors searched records contained in the Safety Board's Aviation Accident/Incident Database, looking specifically for low-airspeed events in the approach and landing phases of flight where equipment failure was not cited as a contributing factor. This search identified 40 low airspeed-related events since 1982. It is likely that additional cases of hard landings and tail strikes have occurred but gone unreported because they did not result in substantial damage. The events identified were categorized by type of operation (Part 121 versus Part 135) and by involvement of structural icing (icing versus non-icing). The results of this categorization are shown in Figure 1. This categorization indicates that low-airspeed events during approach and landing occurred more often during Part 135 than Part 121 flight operations. The results also underscore the prevalence of structural icing in such events. However, that at least 19 of the low-airspeed related accidents and incidents identified did not involve icing or equipment failure.

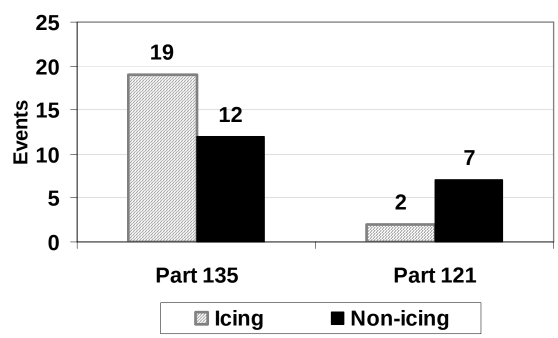


Figure 1. *Accidents and incidents during approach or landing citing low airspeed, 1982-2004.*

Most of the low-airspeed related non-icing events involving Part 121 flight operations resulted in hard landings and/or tail strikes causing substantial aircraft damage, and none resulted in serious injuries. During one typical incident, reported in 1996, the Part 121 airline captain of a McDonnell-Douglas MD-88 said he flew a normal, stabilized approach, using normal flaps and a landing reference speed of 133 knots plus 5 knots. He reported flaring the airplane over the runway and realizing that the sink rate was not being arrested as desired. The captain said he made a more "aggressive" pull on the control yoke while advancing the thrust levers. The airplane landed hard, sustaining substantial damage. Digital flight data recorder readout disclosed that airspeed remained above 138 knots until, at an absolute altitude of 238 feet, airspeed began steadily decreasing below that speed. When the airplane touched down on the runway, airspeed was 125 knots and pitch attitude was 10.6 degrees nose up. There was a +5.5 G vertical acceleration spike at touchdown (Safety Board No. FTW96LA111).

By contrast, low-airspeed related non-icing events involving Part 135 flight operations resulted in more severe outcomes. Records of the investigations of these events indicate that fatal injuries occurred in approximately 1 out of every 4 cases. Part 135 flight operations typically utilize smaller aircraft with less sophisticated autoflight systems. They are less likely to be equipped with auto throttles or sophisticated envelope protection systems. Also, Part 135 flight crews are often less experienced than Part 121 flight crews, and Part 135 flight operations have less stringent flight crew training requirements. These factors could explain the higher prevalence of such events in Part 135 flight operations, and the relative severity of their outcomes.

The Safety Board investigated an accident in 1994, involving a Jetstream 41 on a scheduled Part 135 commuter flight, which crashed 1.2 nautical miles short of the runway during an ILS approach to the Port Columbus International Airport, Columbus Ohio, killing 5 and injuring 2 on board. The flight crew initiated the landing checklist late in the approach. The delay caused distractions to both pilots, and the approach was unstabilized. The autopilot was engaged during the approach, and it kept the airplane on the localizer and glide slope. However, power was set too low to maintain airspeed. This airplane was not equipped with autothrottles. The flight crew did not adequately monitor airspeed indications, and the airplane decelerated until it stalled. Although a stall warning was heard, the captain failed to execute appropriate stall recovery procedures, and the airplane descended steeply, impacting a building. Icing was found not to have been a factor in the accident. The Board found the probable cause of this accident to be, in part, "an aerodynamic stall that occurred when the flight crew allowed the airspeed to decay to stall speed following a very poorly planned and executed approach characterized by an absence of procedural discipline" (National Transportation Safety Board, 1994).

A Change in Design Philosophy

The introduction of a new kind of low-airspeed alert associated with the minimum operationally acceptable speed for a particular phase of flight could help flight crews maintain airspeed awareness in much the same way that altitude alert systems now help flight crews maintain altitude awareness. Such a system would provide an earlier cue to flight crews about low and decreasing airspeed prior to the occurrence of a stall warning, providing them with more time to manage a potential problem before it becomes an emergency.

Recommending a requirement for this kind of low-airspeed alert system represents a departure from the previously accepted premise that adequate low-airspeed awareness is provided by flight crew vigilance and existing stall warnings. However, the history of accidents involving flight crew lack of low-airspeed awareness suggests that flight crew vigilance and existing stall warnings are inadequate to prevent hazardous low-airspeed situations. Moreover, the accident record suggests that this safety issue is not limited to autopilot operations or flight in icing conditions.

The introduction of a low airspeed alerting system could prevent low airspeed / stall related accidents. If a low-airspeed alert had been installed on the King Air involved in the Eveleth accident and had activated when airspeed dropped below 1.2 V_S (about 92 knots), the flight crew could have received about 15 seconds advance warning before the airplane decelerated to its stalling speed. This might have directed the crew's attention to the airplane's decaying airspeed in time to initiate appropriate corrective action. Moreover, if such a system could helped the crew maintain airspeed at or above a minimum operational thresholds such as 1.2 V_S, the likelihood of an accelerated stall initiated by abrupt last-second maneuvering could have been reduced, and improved margins above stalling speed during flight under icing conditions could have been more reliably maintained.

The nature of the airspeed monitoring task varies depending on the level of automation in an airplane cockpit. During a manually flown, a pilot is actively engaged in balancing airspeed, pitch, power, and vertical speed in closed-loop fashion. This requires frequent checking of the outside visual picture and the flight instruments to guide control movements. Alternatively, a pilot using the fully integrated autoflight system in a modern transport airplane monitors flight parameters, including airspeed, in a more supervisory fashion. The issue of airspeed awareness for crews using highly automated flight management systems was raised in an FAA Human Factors Team Report (Federal Aviation

Administration, 1996). Expressing concern about a history of accidents involving lack of low-airspeed awareness among flight crews monitoring automated systems, the report stated:

> Transport category airplanes are required to have adequate warnings of an impending stall, but at this point the airplane may already be in a potentially hazardous low energy state. Better awareness is needed of energy state trends such that flight crews are alerted prior to reaching a potentially hazardous low energy state.

The need for better low airspeed protection and alerting was also cited by the FAA's Flight Guidance System (FGS) Harmonization Working Group of the Aviation Rulemaking Advisory Committee, when, in March 2002, it proposed revisions to 14 C.F.R. Part 25.1329 and associated Advisory Circular 25.1329 to require low-airspeed protection and alerting during autopilot operations for newly certified transport-category airplanes. The proposal stated:

> The requirement for speed protection is based on the premise that reliance on flight crew attentiveness to airspeed indications, alone, during FGS…operation is not adequate to avoid unacceptable speed excursions outside the speed range of the normal flight envelope….Standard stall warning and high speed alerts are not always timely enough for the flight crew to intervene to prevent unacceptable speed excursions during FGS operation….A low speed alert and a transition to the speed protection mode at approximately 1.2 V_S, or an equivalent speed defined in terms of V_{sr}, for the landing flap configuration has been found to be acceptable.

The changes proposed for Part 25.1329 were aimed at future transport category aircraft. However, it may be feasible to develop low airspeed alert systems for less sophisticated, existing airplanes as well. Moreover, the FAA's work, in combination with the Safety Board's accident and incident findings, suggest a need for low airspeed alerting throughout a variety of aircraft with a range of automated features. A low airspeed alert was recently developed for Embraer EMB-120 turboprop airplanes for use in icing conditions. This low airspeed alert system activates an amber-colored indicator light installed in the control panel and provides an auditory alert when airspeed drops below the minimum operational icing speed. In addition, several avionics manufacturers offer low airspeed alerting devices for use in a broad array of general aviation airplanes. These developments suggest that it may be feasible to develop low airspeed alerting systems for most airplane types.

In a letter to the Safety Board dated April 12, 2004, the FAA said it would study cases involving low airspeed awareness that had been identified by the Safety Board and determine what action should be taken. The FAA described existing requirements for stick shakers and stall warnings in transport category airplanes, and cited the increasing prevalence of color-coded visual displays of airspeed found in many modern cockpits. The FAA also stated that it would consider addressing the issue of low airspeed awareness in efforts in progress under its Safer Skies programs and other initiatives. However, as of February 2005, the FAA had not yet announced activities specifically aimed at addressing this issue.

Human Factors Concerns

Technical, operational, and human factors issues must be carefully evaluated and addressed in connection with the design and implementation of any new cockpit alerting system (Pritchett, 2001). Some issues that deserve consideration in association with the possible introduction of new low airspeed alerting systems include: the integration of this system with other aircraft systems; the determination of appropriate threshold speeds for alert activation; examination of the impact of the system's reliability on flight crew confidence in the system; the selection of appropriate strategies for differentiating the alert from existing cockpit alerts and warnings; the development of appropriate flight crew procedures for use in conjunction with the system; and the need for flight crew training in use of the system and related procedures.

Clearly there are many concerns associated with the possible introduction of these systems in commercial airplanes. Despite these concerns, it is possible such systems could significantly improve flight crew performance and increase safety. This is a matter the aviation psychology community is well suited to address. Moreover, the aviation psychology research community has a long history of suggesting and evaluating alternative design solutions for new aircraft systems through applied research.

NTSB Case Numbers for Low Airspeed / Stall Related Events During Approach and Landing[3]

Part 135 Icing Related
SEA82DA017
MKC89LA073
MKC85LA028
MKC84FA033
LAX02LA030
LAX02FA108
DEN90FA068
DEN83LA029
DEN04MA015
DEN01FA094
DCA97MA017
DCA90MA011
CHI98LA084
CHI98FA119
CHI95LA053
CHI86LA090
CHI85FA139
ANC89LA025
ANC02FA020

Part 135 Non Icing Related
MIA89LA193
DEN99FA137
DEN87FA042
DCA94MA027
DCA03MA008
CHI99LA078
CHI01LA109
ANC94LA031
ANC94LA021
ANC91LA015
ANC89LA039
ANC01LA053

Part 121 Icing Related
DCA87IA015
CHI90IA106

Part 121 Non Icing Related
NYC02LA013
LAX90FA148
LAX00LA192
FTW96LA111
BFO85IA036
ATL93IA135
ATL01A064

[3] Information on these cases can be found at http://www.ntsb.gov/ntsb/query.asp.

References

Federal Aviation Administration (2004). FAA-H-8083-3A Airplane Flying Handbook. Washington, DC: U.S. Department of Transportation.

Federal Aviation Administration (1996). Federal Aviation Administration human factors team report on the interfaces between flightcrews and modern flight deck systems. Washington, DC: Author.

Funk, K. (1991). Cockpit task management: Preliminary definitions, normative theory, error taxonomy, and design recommendations. International Journal of Aviation Psychology, 1, 271-285.

National Transportation Safety Board. (2003). Loss of control and impact with terrain, Aviation Charter, Inc., Raytheon (Beechcraft) King Air A100, N41BE, Eveleth, Minnesota, October 25, 2002. Aircraft Accident Report No. AAR-03-03.

National Transportation Safety Board. (1998). In-flight icing encounter and uncontrolled collision with terrain, Comair Flight 3272, Embraer EMB-120RT, N265CA, Monroe, Michigan, January 9, 1997. Aircraft Accident Report No. AAR-98-04.

National Transportation Safety Board. (1994). Stall and loss of control on final approach, Atlantic Coast Airlines, Inc. / United Express Flight 6291, Jetstream 4101, N304UE, Columbus Ohio, January 7, 1994. Aircraft Accident Report No. AAR-94-07.

Pritchett, A. (2001). Reviewing the role of cockpit alerting systems. Human Factors and Aerospace Safety, 1, 5-38.

Ruffel Smith, H.P. (1979). A simulator study of the interaction of pilot workload with errors, vigilance and decisions. NASA Technical Memorandum 78482.

Vakil, S. (2000). Analysis of Complexity Evolution Management and Human Performance Issues in Commercial Aircraft Automation Systems. Unpublished doctoral dissertation. Massachusetts Institute of Technology.

Wickens, C.D. (2003). Pilot actions and tasks: Selections, execution, and control. In P.S. Tsang & M.A. Vidulich (Eds.) Principles and Practice of Aviation Psychology. Mahwah, New Jersey: Erlbaum.

ATCS AGE AND EN ROUTE OPERATIONAL ERRORS: A RE-INVESTIGATION

Dana Broach
FAA Civil Aerospace Medical Institute (AAM-520)
P.O. Box 25082, Oklahoma City, OK 73125

Public Law 92-297 requires that air traffic control specialists (ATCSs), hired on or after May 16, 1972, retire at age 56. This law is based on testimony given in 1971 that as controllers aged, the cumulative effects of stress, fatigue (from shift work), and age-related cognitive changes created a safety risk (U.S. House of Representatives, 1971). The hypothesis has been considered in two studies of en route operational errors (OEs) with contradictory results (Center for Naval Analyses Corporation (CNAC), 1995; Broach, 1999). The purpose of this re-investigation was to test the hypothesis that controller age, controlling for experience, was related to the occurrence of OEs using a statistical method appropriate for rare events. A total of 3,054 usable en route OE records were extracted from the FAA OE database for the period FY1997 through FY2003 and matched with air route traffic control center (ARTCC) non-supervisory controller staffing records, resulting in a database of 51,898 records. Poisson regression was used to model OE count as a function of the explanatory variables age and experience using the SPSS® version 11.5 General Loglinear (GENLOG) procedure. The Poisson regression model fit the data poorly (Likelihood Ratio $\chi^2 = 283.81$, $p < .001$). The odds of OE involvement, estimated with the Generalized Log Odds Ratio, for older controllers (GE age 56) were 1.02 times greater than the odds for younger (LE age 55) controllers, with a 95% confidence interval of 0.42 to 1.64. The range of odds indicated that neither age group was less or more likely to be involved in an OE, controlling for experience. This analysis does *not* support the hypothesis that older en route controllers are at greater risk of involvement in an OE. This finding suggests that the original rationale for the mandatory retirement of ATCSs may need to be re-evaluated. Additional research is recommended.

Public Law 92-297 requires that Air Traffic Control Specialists (ATCSs), hired on or after May 16, 1972 by the FAA, retire at age 56[1]. Controllers with “exceptional skills and abilities” may be given a waiver and continue working until reaching the 61st birthday. The primary evidence offered in support of the mandatory retirement of ATCSs at age 56 in 1971 consisted of anecdotal reports of stress from controllers, studies of self-reported “stress-related” symptoms, physiological correlates of stress, and medical disability retirements of controllers. Despite strong assertions made by various parties, no testimony or data were presented in 1971 to demonstrate that older controllers were more likely than younger controllers to make errors that might compromise the safety of flight.

Several studies of ATCS age and performance have been conducted since passage of P.L. 92-297 (see Broach & Schroeder, in press, for a review). A variety of measures of job performance have been examined in research, ranging from over-the-shoulder subjective evaluations to computer-based measures. Three studies focused specifically on operational errors (OEs). An OE results when an ATCS fails to maintain appropriate separation between aircraft, terrain, and other obstacles to safe flight. OEs are rare compared to the number of operations handled in the U.S. air traffic system. For example, there were 1,145 OEs in fiscal year (FY) 2000 compared to 166,669,557 operations, or 6.8 OEs per million operations (Pounds & Ferrante, 2003; DOT Inspector General, 2003a). Despite their rarity, OEs may pose safety risks, depending on the degree to which separation is lost, and are critical safety indicators for the operation of the air traffic control system (Department of Transportation Inspector General, 2003a,b). OEs occur when through a controller’s actions (or inaction), less than standard separation is maintained.

Spahn (1977) investigated the relationship of age to System Errors (now called Operational Errors) and concluded that “no age group has neither more nor less than its proportional share of system errors” (p. 3-35). The Center for Naval Analyses Corporation (CNAC) found in 1995 that the likelihood of an OE in the period January 1991 to July 1995 declined dramatically in the first few years at an air route traffic control center (ARTCC) and then appeared to approach a constant value. However, CNAC did not examine controller age nor control for age effects. Broach (1999) re-analyzed the CNAC data set from the perspective of controller age and found that the likelihood of an OE might increase with age. The regression analysis also found that experience might mitigate the risk of an OE associated with increasing age. Additional research on the relationship of chronological age, experience, and OEs was recommended. The present study builds on that recommendation. This study was designed to test the hypothesis that older controllers were more likely than younger controllers to commit errors that reduced the safety of flight.

Method

Source Data

A total of 3,054 usable en route OE records were extracted from the FAA Operational Error/Deviation System (OEDS) for the period FY1997 through FY2003. Records for controllers employed at ARTCCs were extracted from the FAA Consolidated Personnel Management Information System (CPMIS) for each fiscal year. There was one CPMIS record in a year for each controller. The OE and CPMIS records were matched by controller identifier and year, producing a database with 51,898 matched records. The number of ATCS with and without OEs is presented by fiscal year in Table 1. For example, of the 7,178 non-supervisory ATCS stationed at ARTCCs in FY1997, 6,864 (95.6%) had no operational errors, while 303 controllers (4.2%) had one OE, and 11 had 2 errors (0.2%). No ATCS had 3 errors in that fiscal year.

Methodological Considerations

Both CNAC (1995) and Broach (1999) calculated the dependent variable of interest as the ratio of controllers with errors in an experience or age range to the total number of controllers in that experience or age range. CNAC labeled this ratio as the "likelihood" of involvement in an error. In fact, both CNAC and Broach calculated the proportion of controllers in a given category that were involved in an error at a given point in time, that is, the prevalence rate. The result is a person-based estimate of risk. However, a person-based estimate of risk does not take into account the varying degrees of exposure between controllers. For example, a controller working a busy, low-altitude transitional sector with multiple merging airways that feed a major hub during an afternoon rush will have a greater opportunity to commit an OE than another controller working a high-altitude sector with sparse cross-continental traffic in steady, predictable east/west flows. Time on position may vary as well. For example, a controller working longer on a given position will have greater opportunity to commit an OE than another controller working less time on a position. As noted by Della Rocco, Cruz, and Clemens (1999), a measure of exposure is required to analyze the risk of being involved in an OE appropriately. However, such measures were unavailable for the present study, leaving the count of errors and prevalence as the variables of interest.

Analysis of counts, such as the number of OEs committed by a controller during a specified period of time, poses analytic challenges. Events such as OEs are rare, compared to the number of operations in the air traffic control system, the number of hours worked by controllers, or even the number of controllers working. While rare events such as OEs are important because of their signal value and potential costs, they are also difficult to study (Hulin & Rousseau, 1980). Techniques borrowed from epidemiology such as count-oriented regression have proven useful in the analysis of rare events. Poisson regression, a count-oriented regression technique, was used in the present study to investigate the degree to which the number of errors is related to controller age.

Poisson Regression

Poisson regression is a statistical technique used to model the expected count of some event as a function of one or more explanatory variables. Examples of events that follow a Poisson distribution are doctor visits, absenteeism in the workplace, mortgage pre-payments and loan defaults, bank failures, insurance claims, and airplane accidents (Cameron & Trivedi, p. 11). In statistics, the "law of rare events" states that the total number of events of interest will take, approximately, the Poisson distribution if (a) the event may occur in any of a large number of trials, but (b) the probability of occurrence in any given trial is small (Cameron & Trivedi, 1998). This statistical "law of rare events" might apply to air traffic control operations as well: there are a large number of aircraft under the control of a relatively large number of controllers at any given moment, but the likelihood of an OE for any given aircraft by any single controller is very small. In this application, the analytic goal was to model the number of OEs incurred by a controller as a function of age and experience (e.g., tenure in the FAA).

Procedure

The data for this analysis consisted of the 51,899 records for non-supervisory center controllers with and without OEs for the period FY1997 through FY2003 (see Table 1). Tenure was recoded into discrete categories to simplify the analysis. The first category for tenure was based on the average of about three years required to complete on-the-job training for center controllers (Manning, 1998). The next interval was 6-years wide (4 through 9), followed five-year increments (Table 2). Age was recoded into two groups: age 55 and younger; and age 56 and older. This split was used to specifically assess the risk that might be associated with controllers older than the mandatory separation age.

Table 1: *N non-supervisory en route ATCS on-board with 0, 1, 2, or 3 operational errors by fiscal year*

	N ATCS with Operational Errors (OEs)				
Fiscal Year	0	1	2	3	AOB Total
1997	6,864	303	11	0	7,178
1998	6,932	389	16	0	7,337
1999	6,869	422	21	0	7,312
2000	6,833	487	31	0	7,351
2001	6,827	549	45	1	7,422
2002	7,110	416	32	0	7,558
2003	7,410	313	17	1	7,741

Table 2: *Tenure by age cross-classification table for Poisson regression analysis*

	Number of OEs (n_{ij})		ATCS Population (N_{ij})	
Tenure Group	LE Age 55	GE Age 56	LE Age 55	GE Age 56
LE 3 Years	44	4	3,587	110
4 – 9 Years	488	10	7,574	191
10 – 14 Years	1,112	20	15,758	280
15 – 19 Years	1,007	2	14,816	128
20 – 24 Years	343	2	5,615	67
GE 25 Years	142	57	2,587	1,186

The data were aggregated by fiscal year, age group, and tenure group to create a cross-classification table suitable for Poisson regression, as shown in Table 2. The columns labeled "Number of OEs (n_{ij})" contain the counts of OEs reported for each age and tenure group combination. For example, there were 44 OEs in the period FY1997 to FY2003 for controllers age 55 or less and with 3 years or less tenure, and 4 OEs for controllers age 56 or older and with 3 years or less tenure. The columns labels "ATCS Population (N_{ij})" contain data representing the number of controllers "exposed" to the risk of incurring an OE during the observation period for each age-tenure combination. For example, there were 3,587 records for en route controllers age 55 or less with 3 years or less tenure who were "at risk" of incurring an OE during the observation period. The goal of the regression analysis is to assess the relative effects of age and tenure on the ratios of errors to "at risk" population. The SPSS® version 11.5 General Loglinear (GENLOG; SPSS, 1999) method was used to conduct the Poisson regression analysis

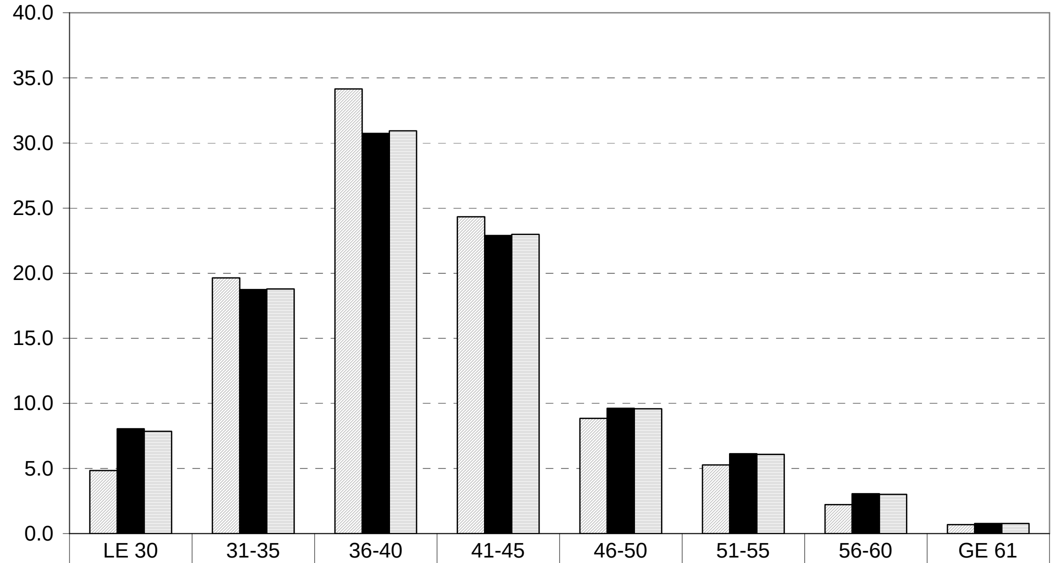
Results

Descriptive Statistics

The initial analyses consisted of simple descriptive statistics. First, the number of OEs per age group for the observation period (FY1997 through 2003) was examined, as shown in Table 2. In this analysis, each controller could have as many as seven records, one for each fiscal year. The records were pooled and then broken out by the number of OEs reported for that age group across the 7 years of observation. As shown in Tables 2, most controllers were not involved in an operational error during the 7-year period. Moreover, the error distribution appears to be similar to the distribution of age, that is, more errors are observed for the more populous age groups. The distribution of controllers with no and one or more OEs by age group is illustrated in Figure 1, relative to the age distribution for all non-supervisory enroute controllers. As found by Spahn in 1977, the distribution of errors by age was very similar to the distribution of age across controllers. No particular age group appeared to experience OEs at a rate disproportionate to their representation in the workforce.

Poisson Regression

Overall, the Poisson regression model fit the data poorly (Likelihood Ratio $\chi 2 = 283.81$, $p < .001$). The parameter estimate for the main effect of age (3.50) was significantly different from 0 (with a 95% confidence interval of 3.29 – 3.70), as were the parameter estimates for tenure. To consider the effect of age across tenure, the two age groups were contrasted. The Generalized Log-Odds Ratio was



	LE 30	31-35	36-40	41-45	46-50	51-55	56-60	GE 61
OE>=1	4.8	19.6	34.2	24.3	8.8	5.3	2.2	0.7
OE=0	8.0	18.7	30.7	22.9	9.6	6.1	3.1	0.8
Total ATCS	7.9	18.8	30.9	23.0	9.6	6.1	3.0	0.8

Figure 1: *OE Involvement by age group compared to distribution of age for all ARTCC controllers, FY1997-2003*

used to estimate the odds ratio for age, that is, the odds of OE involvement for older (GE age 56) controllers (see SPSS, 1999, p. 202 – 203). The odds of OE involvement for older controllers (GE age 56) were 1.02 times greater than the odds for younger (LE age 55) controllers, with a 95% confidence interval of 0.42 to 1.64. A confidence interval for the odds ratio that includes 1.0 indicates that the odds of involvement for the two groups are equal: neither age group was less or more likely to be involved in an OE.

Discussion

The Poisson regression analysis did not support the hypothesis that the likelihood of involvement in an en route OE increased with age. This finding undermines the explicit assertion that early retirement of controllers was "primarily a safety measure" (Testimony of Donald Francke, U.S. House of Representatives, 1971). As noted by Li, Baker, Grabowski, Qiang, McCarthy and Rebok (2003), age in and of itself may have little bearing on safety-related outcomes if factors such as individual job experience, workload, traffic complexity, and time-on-position are taken into consideration (p. 878). For example, supervisors may assign older controllers to less difficult sectors or provide assign an assistant controller during periods of heavy traffic. All other things being equal, age may influence performance through two conflicting pathways. On the one hand, the inevitable changes in cognitive function, particularly speed of processing, may result in slower and less efficient performance. On the other hand, experience is gained with age, and compensatory strategies and meta-strategies may result in safer and more efficient performance by controllers. Additional research on OEs, age, and ATCS performance is recommended to extend and confirm the findings of the present study.

References

Broach, D. (1999, May). *An examination of the relationship between air traffic controller age and en route operational errors.* Paper presented at the 10th International Symposium on Aviation Psychology, Columbus, OH.

Broach, D. & Schroeder, D. (In press). *ATCS Age 56 Report 1: Review of the Scientific Basis for the Mandatory Separation of an Air Traffic Control Specialist (ATCS) at Age 56.* Washington, DC: Federal Aviation Administration Office of Aerospace Medicine.

Cameron, A. C. & Trivedi, P. K. (1998). *Regression analysis of count data.* New York: Cambridge University Press.

Center for Naval Analyses Corporation. (1995, November). *Analysis of operational errors for air route traffic control centers. Volume I: Final report.* (IPR 95-5108). Alexandria, VA: The Center for Naval Analyses Corporation Institute for Public Research.

Della Rocco, P. S., Cruz, C. E., & Clemens, J. A. (1999). Operational errors/deviations and shift work in air traffic control. In Della Rocco, P.S., (Ed.). *The role of shift work and fatigue in air traffic control operational errors and incidents* (pp. 11–21). (DOT/FAA/AM-99/2). Washington, DC: Federal Aviation Administration Office of Aerospace Medicine.

Department of Transportation Inspector General. (2003a). *Operational errors and runway incursions: Progress made, but the number of incidents is still high and presents serious safety risks.* (DOT OIG AV-2003-040). Washington, DC: Department of Transportation Office of the Inspector General.

Department of Transportation Inspector General. (2003b). *DOT's Top management challenges.* (DOT OIG PT-2004-006). Washington, DC: Department of Transportation Office of the Inspector General.

Hulin, C. L. & Rousseau, D. M. (1980). Analyzing infrequent events: Once you find them, your troubles begin. In K. Roberts & L. Burstein (Eds.), *New directions for methodology of social and behavioral science 6: Issues in aggregation* (pp. 65–75). San Francisco, CA: Jossey-Bass.

Li, G., Baker, S. P., Grabowski, J. G., Qiang, Y., McCarthy, M. L., & Rebok, G. W. (2003). Age, flight experience, and risk of crash involvement in a cohort of professional pilots. *American Journal of Epidemiology*, *157*, 874 – 880.

Manning, C. A. (1998). Air traffic control specialist field training programs, 1981-1992. In D. Broach (Ed.). *Recovery of the FAA Air traffic control specialist workforce, 1981-1992* (pp. 23–32). Washington, DC: Federal Aviation Administration Office of Aviation Medicine.

Pounds, J. & Ferrante, A. S. (2003). *FAA Strategies for reducing operational error causal factors.* (DOT/FAA/AM-03/19). Washington, DC: Federal Aviation Administration Office of Aerospace Medicine.

Spahn, M. J. (1977). *The human element in air traffic control: Observations and analyses of the performance of controllers and supervisors in providing ATC separation services. Supplement 2: Analysis of the System Effectiveness Information System (SEIS) data base.* (MITRE Report MTR-7655 Supplement 2). McLean, VA: The MITRE Corporation.

SPSS, Inc. (1999). *SPSS Advanced models 10.0.* Chicago, IL: Author.

Tsang, P. S. (2003). Assessing cognitive aging in piloting. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practice of aviation psychology* (pp. 507–546). Mahway, NJ: Erlbaum.

U.S. House of Representatives. (June 8 & 9, 1971). *Career program for air traffic controllers: Hearings before the Committee on Post Office and Civil Service on H.R. 8083 and Similar Bills.* (House of Representatives, Ninety-Second Congress, First Session. Serial No. 92-10). Washington, DC: Government Printing Office.

End notes:

[1]Mandatory separation is not required for controllers hired before May 16, 1972. The number of controllers age 56 and older increased from 155 in FY1997 to 488 in FY2003.

ANALYZING THE PHYSICAL AND VESTIBULAR EFFECTS OF VARYING LEVELS OF IMMERSIVE DISPLAYS FOR CONTROLLING UNMANNED AERIAL VEHICLES FROM AN AIRCRAFT PLATFORM

John Bruyere, **Justin Gripp**, **Christopher Nagy**, and **Terence Andre**
USAF Academy
Colorado Springs, Colorado

This study attempted to further the base of knowledge concerning effects on watching video taken from an Unmanned Aerial Vehicle (UAV). Sixteen participants from the U.S. Air Force Academy were involved in watching UAV video under 2 conditions of motion (with and without) and 2 conditions of video presentation (laptop computer screen and a head-mounted display). Each video was about 5 minutes long and following each condition the subject filled out a questionnaire which judged their sickness level based on many different factors. Our results did not show any significant difference in sickness levels between the 4 conditions, and further research will have to be performed to fully investigate the effects of watching UAV camera video in an immersive environment.

Introduction

The military has significantly increased its focus on Unmanned Aerial Vehicles (UAVs) in recent years and as this focus and need increases, increased applications of UAVs will continue to arise. While most UAVs have been controlled from a ground station so far, the near future will likely present the need to control UAVs from mobile systems on the ground and military aircraft deployed to an area of interest. Controlling UAVs from other airborne vehicles presents some unique challenges. In particular, operators will have to deal with the potential of sensory conflict between the display from the UAV workstation and the sensory input from the motion of the aircraft. In anticipation of this future need, understanding the unique demands that this will place on the operator(s) of the UAV is crucial. Furthermore, determining which type of display mode will be important when trying to control UAVs from airborne platforms. Consideration should be given to portable 3-D immersive displays (e.g., Head-Mounted Display, HMD) as well as a 2-D laptop computer screen (LCS) in presenting UAV information.

Directly applicable research in this new and focused area is scarce, but there has been a fair amount of research on the slightly broader areas of motion sickness and effects of virtual reality environments. The main theory behind the origins of motion sickness in different environments is the sensory-conflict theory (Yardley, 1992). "Sensory-conflict theory proposes that symptoms occur as a result of conflict between signals received by the three major spatial senses: the visual system, the vestibular system, and nonvestibular proprioception" (Cobb, Nichols, Ramsey, & Wilson, 1999, p. 170). In their study, Cobb et al. (1999) analyzed nine different experiments examining after-effects from different virtual reality (VR) systems, virtual environment (VE) designs, and task requirements, resulting in a total sample pool of 148 participants. A variety of measures, from surveys to physiological indicators, were used to measure different effects, sickness being the item we are most interested in (Cobb et al., 1999). Their results from the self-report data indicated that symptoms of sickness usually occurred within 15 minutes of being immersed. They also found that symptom levels were highest on the first immersion trial and negligible on the third, leading to the conclusion that the participants habituated to the environment after two trials. This information would be helpful in designing our experiment.

As any person experienced with flight simulators knows, UAV pilots today and those of the future will likely have to deal with the phenomenon of vection. Vection "refers to the powerful illusory sensation of self-motion induced in viewing optical flow patterns" (Hettinger, Berbaum, Kennedy, Dunlap, & Margaret, 1990, p. 172). Hettinger et al. (1990) performed a study in which subjects sat and watched a 15-minute flight simulation video which included turns, banks, and altitude changes. The subjects had to watch the display and indicate how much vection was experienced. The researchers hypothesized that those who had more experiences of vection would be more likely to experience sickness. Hettinger et al. (1990) found that those who experienced vection got sick a significantly higher percentage of the time than did those with very limited or no experiences of vection, showing that symptoms of motion sickness can arise by just viewing a screen, screens that usually involve a large field-of-view, even when there is no physical movement. This finding is very relevant to our condition using the HMD because it will likely have a large field-of-view while the individual is tracking an object, compared to our LCS condition, helping us to determine which viewing method is best.

Hettinger et al. (1990) also explained that very few subjects reported symptoms following the initial 15-minute display because motion sickness is a cumulative phenomenon. An important research question in our study is whether a small change in presentation mode (LCS vs. HMD) can impact the feeling of sickness in just a short amount of time.

Studying more about the effects of virtual environments (VE), Stanney, Kingdon, Graeber, and Kennedy (2002) performed a study in which individuals were exposed to a 3-D VE and required to perform certain tasks, such as locomotion, manipulation, turning, etc. The researchers found that the more movement control VE users had, the more presence they would experience, although complete control would make them sicker (Stanney, et al., 2002). These results indicate that there is something of a tradeoff between full movement control in an environment, leading to a higher sense of presence along with a greater level of sickness, and less control, correlated with less sickness. This could be important for how much immersion and how much control (versus possibly more automation) is presented to pilots controlling UAVs.

Another study that is directly applicable to our experiment was performed by Kennedy, Lane, Berbaum, and Lilienthal (1993) which consisted of creating a new survey from which to judge simulator sickness (SS). The Simulator Sickness Questionnaire (SSQ) was developed to replace the Pensacola Motion Sickness Questionnaire (MSQ) which was not created for SS and thus less applicable to the special circumstances of SS. The SSQ was derived partly from the MSQ using a series of factor analyses to come up with the appropriate format and substance. Our study used a modification of the SSQ in order to compare the levels of sickness experienced by our subjects.

In our experiment, we introduced two independent variables (IVs) with two conditions each. The first IV is whether the students viewed camera video with an immersive HMD or with a non-immersive LCS. The second IV is whether the subject experienced motion in the chair they were sitting in during the trial. Motion is defined as random turns, from 90 degrees to 180 degrees, that the subject experienced. This motion is meant to simulate the motion that might be felt onboard an aircraft, though it is severely limited in that it can only move in two degrees of motion and cannot simulate turbulence. The dependent variable (DV) is simply what level of sickness they feel, defined by their results from the SSQ.

Our null hypothesis is that there will be no significant difference between the HMD and LCS conditions on the level of reported nausea in the participants. Our alternative hypothesis is that the HMD environment will increase the feeling of motion sickness, and that when coupled with movement, the increase will be even greater. We feel that when fully immersed, an individual will feel more sickness than when not immersed, and that those feelings will be intensified with even slight motion.

Method

Participants

For our study, we utilized 16 male and female cadets enrolled in an introductory psychology class at the United States Air Force Academy. These cadets consisted of both freshman and sophomore cadets who voluntarily signed up and received extra credit from their teachers for participating.

Apparatus

In order to expose the participants to a UAV environment, we used a fully immersive HMD virtual reality system to recreate the pilot's view. The HMD was a Virtual Research V8 with 800x600 resolution. To recreate the view from a flat screen display to compare against the HMD, we used an IBM ThinkPad Laptop computer for the non-immersive Laptop Computer Screen (LCS). A Dell 3.2 GHz, Pentium 4 processor desktop computer was used to display the video for the HMD. For the video, we used UAV flight video flown over recognizable areas of the United States Air Force Academy which lasted 4 minutes and 40 seconds in length. To simulate the movements for the participants, we used a non-motorized Barany Chair. A modified SSQ was given to each participant in-between each condition. A picture of the setup can be seen in Figure 1.

Figure 1. *Experiment setup of HMD condition.*

Procedure

Upon arriving to the testing room, the participants read and completed the consent form. After they were told that they may withdraw at anytime, we explained the procedures and began the experiment. Participants watched the UAV camera video in 4 different conditions. Those conditions were the HMD with motion, the HMD with no motion, the LCS with motion, and the LCS with no motion. Our experimental design includes a balanced within-subjects design in which each participant watched the video under all four conditions. The design is balanced so that each possible order of the four conditions is performed in order that learning effects do not affect the outcome of the data. For the movement conditions, the participant began with the chair facing south. The participants were exposed to lateral turns in the Barany chair at exact times and rotation degree values predetermined by the researchers as shown in Table 1.

Table 1. *Times and directions for chair movements during experiment.*

Time	Direction	Degrees
0:15	L	90
0:45	R	90
1:00	R	90
1:20	L	180
1:25	R	180
2:00	L	90
2:10	R	90
2:40	L	180
3:00	R	180
3:10	L	90
3:15	L	90
3:45	R	180
4:00	L	180
4:10	R	90
4:30	L	90
4:35	R	90

For the "no movement" conditions, the participants sat in the Barany chair during the video presentation. For the "movement" conditions, the participants wore the HMD; during the LCS conditions, participants held the laptop in their lap for the duration of the video. For all conditions, the lights were turned off in the room in order to prevent any distractions and to simulate a dark aircraft cabin. All the participants were given tasks to complete during the video under all conditions. They were told to count and keep a running total in their heads of any moving vehicles they saw, count and keep a running total in their heads of any street intersections, and provide heading information (North, South, East, or West) pertaining to the UAV flight path. After each condition, the participant completed a modified version of the SSQ and was given 1 minute to walk around and rest before the next condition began. The modified SSQ we used did not include 3 components that did not apply to our very basic simulator experience or might be confusing based on our pilot study (i.e., stomach awareness, burping, and fullness of head). We replaced two of these terms with more general items (dizziness and sickness feeling) hoping to make the questionnaire slightly more sensitive to the very slight differences that might exist between our conditions.

Results

The data generated by the SSQ was entered into SPSS v 11.0 and analyzed. The data from each condition was matched with each participant based upon the order in which the participants signed up to participate in the study. Throughout the analysis the alpha was selected as $\alpha = .05$.

Participant scores for each item of the SSQ were summed according to a weighted scale subscribed by the SSQ as illustrated in Table 2. Simulator sickness has three subcomponents which make the whole construct (nausea, oculomotor, and disorientation). Certain items such as difficulty focusing and nausea counted towards two of the three subcomponents. Other items, like fatigue, counted toward just one subcomponent. No item counted towards all three. We made an assumption that the new items we replaced on the SSQ (dizziness and sickness feeling) counted in a single category as did the original items. We multiplied any item answer that counted towards two subcomponents by a factor of two to weight it properly and any answer that counted toward just one subcomponent was multiplied by 1. We then summed all of the scores for each condition in order to obtain a total sickness score for each participant in all four conditions.

Table 2. *Weighting scale showing example of weights for two symptoms.*

	Weight		
SSQ Symptom	*Nausea*	*Oculo-motor*	*Disorientation*
Fatigue	0	1	0
Nausea	1	0	1

We analyzed our data using a repeated measures analysis of variance. Table 3 shows the mean and standard deviations for each of the four conditions. For the display presentation type, we found no significant results, $F(1,15) = 3.615$, $p = 0.077$. For motion we again found no significant results, $F(1,15) = 1.90$, $p = 0.188$. There was also no interaction between display type and motion, $F(1,15) = 2.241$, $p = 0.155$. Figure 2 shows a graph of the changes in sickness means for both of our conditions (motion and display).

Table 3. *Mean and standard deviation SSQ scores for all four conditions.*

	Display Type	
	HMD	LCS
Motion Type	Mean (SD)	Mean (SD)
Motion	7.125 (6.20)	3.4375 (3.65)
No Motion	4.8750 (4.60)	3.3750 (3.42)

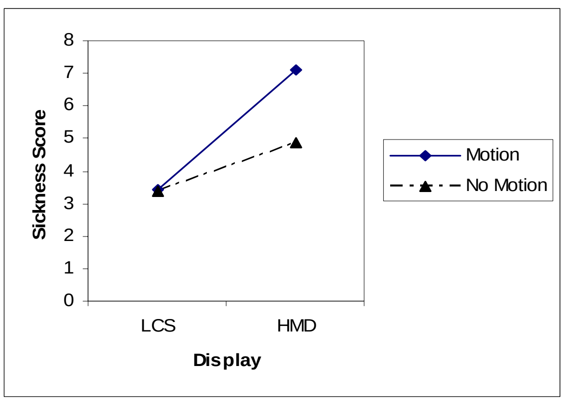


Figure 2. *A graph showing the marginal means for all four conditions.*

Discussion

The mean values for the HMD conditions showed a slight increase in the level of sickness reported, although the difference was not statistically significant using an alpha = .05. There are many probable reasons for why we were unable to show any significant results. The first and most salient reason is the number of participants. Under optimal conditions, we would need approximately 30 participants in each cell to show reliable results. In our case we were limited to 16 based on class time constraints and scope of project. Perhaps with a larger participant pool we would be able to show the results we had expected. A second mitigating factor was our inability to utilize motion on any axis outside of the z-axis. In a real life condition, a UAV pilot operating from an AWACS would be subject to motion on all three axes, not just one. Another real-life condition we did not have the means to replicate was the flying of a UAV. We subjected our participants to videos of UAV flight, but this condition could not fully represent the attention that would be given to the screen if the participant was actually piloting a UAV, even despite our attempts to alleviate this factor by having the participant attend to heading, moving vehicles, and intersections. Another factor that influenced our results was the amount of time the participants were subjected to the video. Due to time constraints, we were only able to show a 4 minute and 40 seconds video, where Cobb et al. (1999), found that 15 minutes of immersion is necessary to generate the sickness we were looking for.

For future research, we would recommend that the above issues be addressed by meeting a few important requirements (assuming optimal conditions and appropriate assets). First the sample size should be increased to include at least 30 participants to help validate the findings. Secondly, a full-motion and fully immersive simulator should be utilized, in which the participants would be required to actually operate a simulated UAV. Finally, participants should be subjected to each condition for at least fifteen minutes before the SSQ is administered.

References

Cobb, S. V., Nichols, S., Ramsey, A., & Wilson, J. R. (1999). Virtual reality-induced symptoms and effects (VRISE). *Presence: Teleoperators and Virtual Environments, 8*(2), 169-187.

Hettinger, L. J., Berbaum, K. S., Kennedy, R. S., Dunlap, W. P., Margaret, D. N. (1990). Vection and simulator sickness. *Military Psychology, 2*(3), 171-181.

Kennedy, R. S. Lane, N. E., Berbaum, K. S., & Lilenthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*. *3*(3), 203-220.

Stanney, K. M., Kelly, S., Graeber, D., & Kennedy, R. S. (2002). Human performance in immersive virtual environments. *Human Performance, 15*(4), 339-367.

Yardley, L. (1992). Motion sickness and perception: a reappraisal of the sensory conflict approach. *British Journal of Psychology, 83*(4) 449-472.

EFFECTS OF WORKLOAD AND LIKELIHOOD INFORMATION ON HUMAN RESPONSE TO ALARM SIGNALS

Ernesto A. Bustamante
James P. Bliss
Old Dominion University
Norfolk, VA

The purpose of this study was to examine how workload and likelihood information would affect participants' responses to alarm signals while they performed a battery of tasks. As expected, participants' overall response rates and false alarm response rates were significantly lower, and true alarm response rates were significantly higher when they used a likelihood alarm system. These results were particularly noticeable under high workload conditions. Results from this study suggest that although people may respond less often to alarm signals when they are provided with likelihood information, they will more likely respond to true signals rather than false alarms. Therefore, designers should incorporate likelihood information in alarm systems to maximize people's ability to differentiate between true and false alarms and respond appropriately.

Introduction

Technological advances have made the use of automated alarm systems a common practice in aviation (Bliss, 2003). Such systems serve a crucial function in the cockpit by alerting pilots of potential or imminent dangerous conditions. Nevertheless, even the most sophisticated alarm systems emit a high number of false alarms, increasing pilots' level of workload and jeopardizing their flight performance (Getty, Swets, Pickett, & Gonthier, 1995; Gilson & Phillips, 1996).

A possible solution to this problem is to provide pilots with additional information regarding the positive predictive value (PPV) of alarm signals through the use of a likelihood display. The PPV of a signal, which is also commonly referred to as its "alarm reliability," is defined as the conditional probability that given an alarm, a problem actually exists. Researchers have shown that people adjust their responsiveness based on the outputs given by alarm systems (Meyer & Ballas, 1997; Robinson & Sorkin, 1985). More specifically, people's responsiveness to alarm signals is dependent on the PPV of such signals (Bliss & Dunn, 2000; Bliss, Gilson, & Deaton, 1995; Getty et al., 1995). The purpose for using a likelihood alarm display is to provide people with information about the PPV of different signals so that they can respond more often to high-likelihood signals and less often to low-likelihood signals.

However, researchers have questioned the usefulness of such displays by pointing out that they may actually decrease pilots' responsiveness, thereby jeopardizing flight safety (Sorkin, Kantowitz, & Kantowitz, 1988). Nonetheless, providing pilots with likelihood information may enhance their decision-making strategies such that they might respond more often to signals that signify actual problems and disregard false alarms. However, few researchers have examined how operators of complex tasks react when faced with signals generated by a likelihood alarm system. Similarly, there is little awareness of how other task variables might interact with likelihood information to influence alarm reaction patterns or primary task performance. The purpose of this study was to examine how workload and likelihood information would affect people' responses to alarm signals.

Participants performed the tracking and resource-management tasks from the Multi-Attribute Task (MAT) Battery (Comstock & Arnegard, 1992) and an engine-monitoring task that the experimenters designed. We manipulated workload level by automating the tracking task and by increasing the difficulty of the resource-management task. While performing their tasks, participants reacted to alarms generated by either a binary alarm system (BAS) or a likelihood-alarm system (LAS).

We assessed participants' response rates to false alarms and true signals. We expected participants to respond more often to false alarms when they interacted with the BAS, particularly during low workload (Sorkin et al., 1988). This hypothesis was consistent with previous research, which suggests that people are generally more likely to respond to alarm signals under low workload conditions (Meyer, 2002). However, we hypothesized that participants would respond more often to true signals when they interacted with the LAS compared to the BAS, and that this difference would be greater under high workload conditions. The reason for this was that we expected the LAS would improve participants' ability to detect alarms that were more likely to be true signals. Such an expectation is reflected by Selcon, Taylor, and Shadrake (1991), who demonstrated the benefits of redundant information on pilot reactions to displays in the cockpit.

Method

Experimental Design

We used a full within-subjects design. Preliminary analyses consisted of descriptive statistics to ensure that we did not violate any statistical assumptions. We set statistical significance for all inferential tests *a priori* at $\alpha = .05$.

Participants

An *a priori* power analysis revealed that approximately 30 participants would be necessary to obtain a power of .80, assuming a medium effect size ($f = .25$) at an alpha level of .05 (Cohen, 1988). Therefore, we used convenience sampling to select 30 (18 females, 12 males) undergraduate and graduate students from Old Dominion University to participate in this study. Participants ranged from 18 to 38 years of age ($M = 22.70$, $SD = 4.54$). All participants had normal or corrected-to-normal vision and hearing. To motivate participants, we provided them with three research credit points to apply to their class grades, and awarded a $10 prize to the person who performed best.

Materials and Apparatus

To increase the realism of the experimental design, participants performed a set of complex primary tasks at the same time they performed the secondary task. The primary tasks consisted of a compensatory-tracking task and a resource-management task, both taken from the MAT (Comstock & Arnegard, 1992). We loaded the MAT on an IBM-compatible computer and displayed it to participants using a 17-inch monitor. Participants performed the MAT using a standard mouse and a QWERTY keyboard.

While performing the MAT tasks, participants also performed an engine-monitoring task that the experimenters designed. We presented this task to participants on a separate 17-inch monitor, located at 90º to the right of the primary task. This engine-monitoring task required participants to respond to a series of alarms that indicated a potential problem with two engines. As they performed the MAT, participants encountered different alarms and had to decide whether to ignore them or respond to them by searching for critical system-status information. To search for this information, participants had to divert their attention from the primary task and press the space bar on the keyboard located in front of the computer hosting the secondary task. Once they did this, the screen presented them with the system-status information regarding the current oil temperature and pressure of the two engines. Participants then assimilated this information and decided whether they needed to correct the problem by pressing the space bar again, or cancel the information by pressing the escape key and returning to the primary task. To keep participants motivated, they received a score on the engine-monitoring task, which was updated after each alarm depending on their response.

Participants received one point for searching for further information when an alarm was true and for ignoring false alarms. They lost one point for searching for further information when an alarm was false, but they lost three points for ignoring a true alarm. If they checked the status of the two engines, they received two points for correctly resetting actual problems and one point for canceling the information when there was no problem. They also lost one point for resetting the system when there was no problem, but they lost three points for canceling the information when a problem actually existed. The rationale for using this point system was to more closely simulate the payoff associated with responding to and ignoring alarm signals in a complex task situation, such as flying an airplane, where adequately responding to true alarms is crucial for flight safety.

Alarm Systems

Binary Alarm System We modeled the performance of the binary alarm system based on prior research (Bustamante, Anderson, & Bliss, 2004). The probability of a problem was .01. The system had a high sensitivity (d'=3.98) and a low threshold (β=.23). Based on these parameters, the system was able to detect the presence of a problem 99% of the time, while issuing a false alarm rate of 5%. The system had a sampling rate of 1s. Each experimental session lasted 30 minutes, and a problem could arise at any given second throughout each session. Based on the prior probability of the problem, a total of 18 engine malfunctions occurred throughout each session. The system was able to detect the presence of all the problems, thereby generating a total of 18 true alarms throughout each session. However, because of the low base rate of the problem and the system's low threshold, it generated a total of 82 false alarms, resulting in an overall system reliability of 18%. The true and false alarms generated by the system looked and sounded exactly alike, to reflect real-world situations where the operator must search for additional information to ascertain alarm validity. The visual component of the alarm signal consisted of a yellow circle accompanied by the word "WARNING" written underneath it. The auditory component of the alarm signal was a simple sine wave at a frequency of 500 Hz, presented at 65 dB(A) through a set of flat-panel speakers. The ambient sound pressure level was approximately 45dB(A).

Likelihood Alarm System. The overall performance of the likelihood alarm system was the same as the binary system. However, this system generated two types of alarms depending on the likelihood that they would be true. To determine the likelihood of each alarm, the system had two simulated thresholds instead of one. We set the lowest threshold of this system at the same value as the binary system, and the highest threshold at β=88.40. Based on these two thresholds, the system generated a total of 84 low-likelihood alarms, 4 of which were true and 80 of which were false. As a result, these alarms had a 5% likelihood of being true. This system generated a total of 16 high-likelihood alarms, 14 of which were true and 2 of which were false. As a result, these alarms had an 88% likelihood of being true. The low-likelihood alarm signals consisted of the same stimuli used for the binary system. The visual component of the high-likelihood alarms consisted of a red circle accompanied by the word "DANGER" written underneath it. The auditory component of these alarms was a simple sine wave at a frequency of 2500 Hz, also presented at 65dB(A).

The rationale for using this particular design for the likelihood alarm system was to use peripheral cues such as color, signal word, and sound frequency to enable participants to easily differentiate between low- and high- likelihood alarms. Although these cues may affect the perceived urgency of such signals, prior research suggests that the effect of the PPV of alarms overshadows any effect that could be attributed to perceived urgency (Burt, Bartolome-Rull, Burdette, & Comstock, 1999).

Procedure

As part of this study, participants completed two experimental sessions during which they interacted with an alarm system and an automatic pilot. During one of these sessions, participants used a binary alarm system, and for the other session, they used a likelihood alarm system. We fully counterbalanced the order in which participants used these systems.

Participants came to the laboratory individually. When they entered the laboratory, they first read and signed an informed consent form and then completed a background information form. The purpose of the background information form was to collect information relevant to the exclusionary criteria for the experiment, such as participants' age and whether they had any visual or auditory problems. Once participants completed this form, we provided them with the instructions about how to perform the MAT tasks. Next, participants performed a 5-min practice session.

Once participants completed this practice session, the experimenter provided them with the instructions about how to complete the engine-monitoring task. Participants then went through another 5-min practice session, performing all tasks at the same time. Next, the experimenter informed participants of the overall reliability of the system and the likelihood of each type of alarm. Then, participants performed the two experimental sessions, taking a 5-min break between them. Before participants began the second session, we provided them with information about the other alarm system. Then, participants went through another 5-min practice session, using the other alarm system. After this practice session was over, participants performed the second experimental session using the other alarm system.

Each experimental session lasted 30 min. During the first and last 7.5 min, participants performed the tracking task manually, and they experienced a series of random pump malfunctions in the resource-management task. At other times, the autopilot performed the tracking task, and participants did not experience any pump malfunctions in the resource-management task. The rationale for doing this was to more closely simulate the distribution of workload levels found in applied settings, such as in aviation, where the take-off and landing phases of flight are associated with higher levels of workload than the cruising phase.

Dependent Measures

We assessed participants' overall response rates (ORR), which was the proportion of alarms that participants responded to in a given session. We also assessed participants' false alarm response rates (FARR), which was the proportion of false alarms that participants responded to in a given session. Last, we assessed participants' true alarm response rate (TARR), which was the proportion of true alarms that participants responded to in a given session.

Results

We conducted three 2 x 2 repeated-measures ANOVAS. We used workload (Low, High) and system (BAS, LAS) as independent variables. We used ORR, FARR, and TARR as dependent measures. Results from the first ANOVA showed a statistically significant main effect of workload on ORR, $F(1,29) = 46.25$, $p < .001$, partial $\eta^2 = .62$. Participants' ORR was significantly higher during low workload ($M = .51$, $SD = .24$) than during high workload ($M = .40$, $SD = .23$). Results from this first analysis also showed a statistically significant main effect of system on ORR, $F(1,29) = 28.04$, $p < .001$, partial $\eta^2 = .49$. Participants' ORR was significantly higher when they interacted with the BAS ($M = .54$, $SD = .26$) than when they interacted with the LAS ($M = .37$, $SD = .19$). These results are shown in Figure 1.

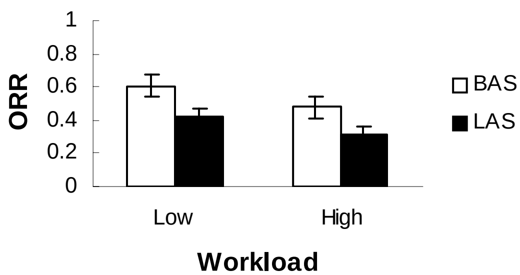


Figure 1. *Overall response rate as a function of workload and system.*

Results from the second ANOVA showed a statistically significant main effect of workload on FARR, $F(1,29)=35.67$, $p<.001$, partial $\eta^2=.55$. Participants' FARR was significantly higher during low workload ($M = .46$, $SD = .27$) than during high workload ($M = .34$, $SD = .26$). Results from this second analysis also showed a statistically significant main effect of system on FARR, $F(1,29)=57.93$, $p<.001$, partial $\eta^2=.67$. Participants' FARR was significantly higher when they interacted with the BAS ($M = .54$, $SD = .25$) than when they interacted with the LAS ($M = .27$, $SD = .22$). These results are shown in Figure 2.

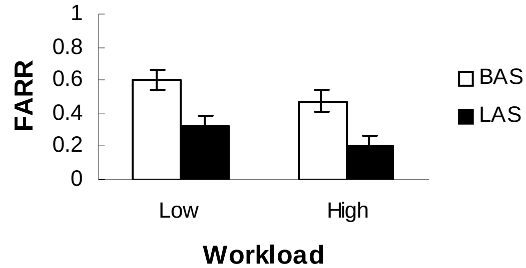


Figure 2. *False alarm response rate as a function of workload and system.*

Last, results from the third ANOVA showed a statistically significant workload by system interaction effect, $F(1,29)=7.20$, $p<.05$, partial $\eta^2=.20$, and statistically significant main effects of workload, $F(1,29)=14.10$, $p<.01$, partial $\eta^2=.33$, and system, $F(1,29)=30.22$, $p<.001$, partial $\eta^2=.51$, on TARR. Participants' TARR was significantly higher when they interacted with the LAS ($M = .80$, $SD = .13$) than when they interacted with the BAS ($M = .56$, $SD = .31$), but this difference was greater during high workload. These results are shown in Figure 3.

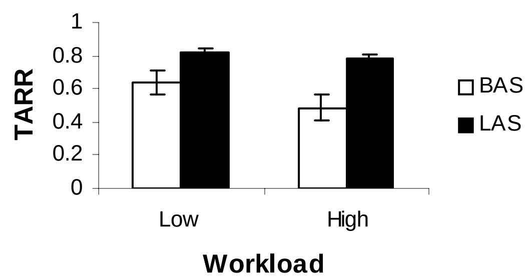


Figure 3. *True alarm response rate as a function of workload and system.*

Discussion

Results supported our hypotheses. As expected, participants responded significantly more often to false alarms when they interacted with the BAS, particularly under low-workload conditions. However, participants responded significantly more often to true signals when they interacted with the LAS, especially during high-workload conditions.

In general, the results of this experiment support the use of redundant information to signify alarm validity, or lack thereof. As noted by Selcon, et al. (1991), the presence of such information can improve pilot reactions to displayed information in the cockpit. Bliss, Jeans, and Prioux (1996) showed similar results; when participants were faced with an unreliable alarm system, they benefited most from the presence of additional information upon which to base their judgments of individual alarm validity.

Results from this study have potential applications for designing alarm systems in the field of aviation. These results suggest that although pilots may respond less often to alarm signals when they are provided with likelihood information, they are more likely to respond to true signals rather than false alarms. Therefore, designers should incorporate likelihood information in alarm systems to maximize pilots' ability to differentiate between true and false alarms and respond appropriately. This, in turn, may increase safety by directing pilots' attention to actual problems without jeopardizing flight performance by minimizing responsiveness to false alarms.

References

Bliss, J. P. (2003). Investigation of alarm-related accidents and incidents in aviation. *The International Journal of Aviation Psychology, 13*(3), 249-268.

Bliss, J. P., & Dunn, M. C. (2000). Behavioural implications of alarm mistrust as a function of task workload. *Ergonomics, 43*(9), 1283-1300.

Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics, 38*(11), 2300-2312.

Bliss, J. P., Jeans, S. M., & Prioux, H. J. (1996). Dual-task performance as a function of individual alarm validity and alarm system reliability information. Proceedings of the IEA 1996/HFES 1996 Congress, Santa Monica, CA: International Ergonomics Association/Human Factors and Ergonomics Society, 1237-1241.

Burt, J. L., Bartolome-Rull, D. S., Burdette, D. W., & Comstock, J. R. (1999). A psychological evaluation of the perceived urgency of auditory warning signals. In N. A. Stanton & J. Edworthy (Eds.), *Human factors in auditory warnings* (pp. 151-169). Aldershot, Hants: Ashgate Publishing Limited.

Bustamante, E. A., Anderson, B. L., & Bliss, J. P. (2004). Effects of varying the threshold of alarm systems and task complexity on human performance and perceived workload. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 1948-1952.

Comstock, J. R., & Arnegard, R. J. (1992). *The multi-attribute task battery for human operator workload and strategic behavior research* (NASA Technical Memorandum No. 104174). Hampton, VA: National Aeronautics and Space Administration, Langley Research Center.

Getty, D. J., Swets, J. A., Pickett, R. M., & Gonthier, D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of Experimental Psychology: Applied, 1*, 19-33.

Gilson, R.D., & Phillips, M.J. (1996). Alert or sound alarm? *Aerospace Engineering*, 21-23.

Meyer, J. (2002). Task demands and responses to warnings. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 308-312.

Meyer, J. & Ballas, E. (1997). A two-detector signal detection analysis of learning to use alarms. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 186-189.

Robinson, D. E. & Sorkin, R. D. (1985). A contingent criterion model of computer assisted detection. In R. E. Eberts & C. G. Eberts (Eds.), *Trends in ergonomics/human factors II* (pp. 75-82). Amsterdam, North-Holland: Elsevier.

Selcon, S.J., Taylor, R.M., & Shadrake, R.A. (1991). Giving the pilot two sources of information: help or hindrance? In E. Farmer (Ed.), *Human Resource Management in Aviation* (p. 139-148). Aldershot: Avebury Technical.

Sorkin, R. D., Kantowitz, B. H., & Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors, 30*(4), 445-459.

HUMAN FACTORS ISSUES OF TCAS: A SIMULATION-BASED STUDY

Cabon, P. ; **Rome, F**. ; **Favresse, A**. ; **Mollard, R.**
Unité d'Ergonomie - LAA University Paris 5 - Paris, France

Figarol, S. ; **Hasquenoph, B**. ; **Houvenagel, C.**
DSNA[1] - SDER- Sub Directorate for Studies and Research **-** Toulouse, France

Since its introduction in the 90's, TCAS II, presented as a straightforward and very reliable technological tool, has significantly reduced the risk of collision. Paradoxically, the introduction of this system has been accompanied with numerous incidents and one major accident in 2002, mainly due to unclear rules, poor air-ground cooperation and poor human decision. In order to investigate these potential human factors issues, a part-task air-ground simulation was conducted: 10 pilots and 10 controllers were involved in the simulations of 4 scenarios containing TCAS occurrences. Data collected included video camera recordings for behavioral analysis, Heart Rate (HR) for stress evaluation, questionnaires and debriefings for perceived risk levels and situational awareness assessment. The observations and errors were analyzed through the CREAM methodology. The debriefings were led through a self-confrontation technique, together with pilots and controllers. Results show that the simulations of TCAS situations were able to produce a significant physiological stress response with significant increase of HR when a resolution happens. Questionnaires and debriefings show that, in most of the observed cases, aircrew, and controllers are not sharing the same mental picture of the involved traffic and the risk of collision. This raises important issues in terms of cooperation between controllers and aircrews in such demanding occurrences. This should allow identifying risky situations and the related generic causes. The results will be discussed, aiming at a potential improvement of the system, in terms of Human Machine Interface, training and consistency of procedures.

Introduction

The prevention of mid-air collision has been a major safety issue in aviation for years. Since its introduction in the 90's, the Traffic Alert and Collision Avoidance System II (TCAS II), presented as a straightforward and very reliable technological tool, has significantly reduced the risk of collision. The latest version, TCAS II Version 7 was built upon lessons learned from TCAS II use and problems (Wickens, 1992) TCAS II is now a mandatory device for all commercial aircraft with more than 19 passengers seats. This system issues two types of alerts : the Traffic Advisory (TA) which identifies a traffic as an intruder whose position should be closely monitored (but no actions are required for the aircrew) and the Resolution Advisory (RA) that recommends a vertical escape maneuver to maintain a self separation. Paradoxically, the introduction of this system has and still contributes to severe incidents and was the main cause of one major accident, the mid-air collision between a B757 and a Tupolev at Uberlingen Lake in 2002. The major cause of this accident lies in the decision of the Tupolev captain to follow, (accordingly to his company's manual), the Air Traffic Controllers (ATC) instruction to immediately initiate a descent though it was contrary to the RA order (BFU, 2004). Even if an improvement seems to show up over the last years mainly due to aircrew and Air Traffic Controllers (ATCO) drastic changes in information and training (Powell and Baldwin, 2002) it is still observed cases where aircrews failed to follow the RA or over-reacted or simply disregarded the alert. Obviously, this system still raises many human factors issues that directly impair air safety. A preliminary study (Cabon et al, 2003) conducted by means of collective and individual interviews of controllers and pilots emphasized the following issues: stress, man-machine interface, training, airline procedures and aircrew-ATC communications. The present study aims to investigate the potential human factors issues in an air-ground simulation. The use of simulation is essential as the previous studies emphasised the need to reproduce in real time the temporal pressure and the stress that experience both pilots and ATCOs during a TCAS sequence.

Method

Simulation Settings

All the simulation settings were designed by the Centre d'Etudes de la Navigation Aérienne (CENA) in Toulouse (France).The three main elements were:

- An Airbus A320 part-task simulator including for both the Pilot Flying (PF) and the Pilot Non Flying (PNF), the main displays and tools that are needed to present and respond to a TCAS resolution: the Navigation Display (ND), the Primary Flight Display

[1] Previous name was CENA (Centre d'Etudes de la Navigation Aérienne), which is part of DSNA.

(PFD), the Flight Control Unit (FCU) and a side stick. Radio communications with ATC are available.

- An ATC position with the 2 radar displays and paper strips for the planning and executive controllers.
- A “pseudo-pilots” position where 2 experts play the role of the surrounding traffic. The ATC did not know during the simulation what aircraft was actually “piloted” or “pseudo-piloted”.

The main and most valuable feature was the integration of the actual TCAS software and HMI in the cockpit simulator and for the other simulated aircraft.

Scenarios

While high technical fidelity was out of scope, operational aspects were taken as important. For this study, 4 scenarios have been especially designed. The first one (Biarritz) was designed by the CENA to induce a high probability to trigger a TCAS alert. In this scenario, always presented first, neither the ATCOs nor the aircrews knew that the study was dealing with TCAS operation. The three other scenarios (named respectively Marseille, Orly and Reims) were based on real incidents that were selected in collaboration with the CENA and the Service du Contrôle du Trafic Aérien (SCTA). In these scenarios, the ATCO were asked to “play a part”, reproducing certain errors in order to induce a conflict likely to trigger a TCAS alert. Each scenario lasted between 10 to 15 minutes.

Participants

A total of 10 A320/330/340 pilots (i.e. 5 aircrews) and 10 ATCOs (ACC and APP) were involved in this study.

At the beginning of each session, none of the participant knew the precise scope of this study, in order to avoid anticipation or preparation effects.

Data Collected

Four kinds of data were collected:

- Direct observations and video of both working positions to trace displays, events, actions and communications to subsequently analyze behavior. Specific observation grids were developed using the Cognitive Reliability and Error Analysis Method (CREAM) (Hollnagel, 1998). On top of this, one of the observer was a fully qualified pilot able to pinpoint fine details not caught by the video. Two Human Factors experts also observed aircrew and ATCO.
- Subjective assessment. After each scenario, participants were asked to fill out questionnaires to rate their situational awareness, their stress and various aspects that were relevant to understand how they had perceived the scenario and the TCAS sequence.
- Heart rate (HR). In order to get an objective measurement of stress, heart rate was continuously recorded during the scenario by means of a digitized recorder (Vitaport, Temec ®).
- Collective debriefing. The aim of the debriefing was to collect the verbalization of both pilots and ATCOs on what happened during the scenarios. The debriefing was supported by an auto-confrontation using the video and communication recordings. This debriefing was very useful to assess the situational awareness of participants. It also allowed revealing their a posteriori understanding of the situation, in relation to the ASR or reports they would have to fill in. At the end, a discussion was set up about the main safety-related issues and suggestions to reduce risk in operational environment.

Each session lasted one day from 0900 to 1730. The four scenarios were played in the morning while the afternoon was dedicated to the collective debriefing.

Results

Descriptive Analysis

TCAS Events

During the study, 20 scenarios have been played (i.e. 4 scenarios X 5 days). Both the simulation setting and the scenarios were efficient to induce a significant number of TCAS events allowing the data analysis. The following TCAS events occurred during the simulations:

- 8 TA not followed by a RA,
- 18 sequences TA/RA (some with several RA),
- 37 RA (initial and sense reversal or weakening RA).

A rather good variability of RA was obtained, with a majority of Adjust Vertical Speed which are known to be often misinterpreted by aircrews.

Heart Rate (HR)

Stress was objectively measured in this study using a continuous recording of HR. As there is a considerable inter-individual variability in HR, all the data are expressed as the percentage of variation of the 1st percentile of the total recording (reference). Figure 1 shows an example of HR recording for a pilot and an ATCO during a TCAS sequence.

This example shows a clear physiological reaction to the occurrence of the different TCAS events for the pilot and the ATCO. For the pilot HR increased dramatically after the TA up to 80% when the two first RA “Climb” and “Adjust Vertical Speed” are issued.

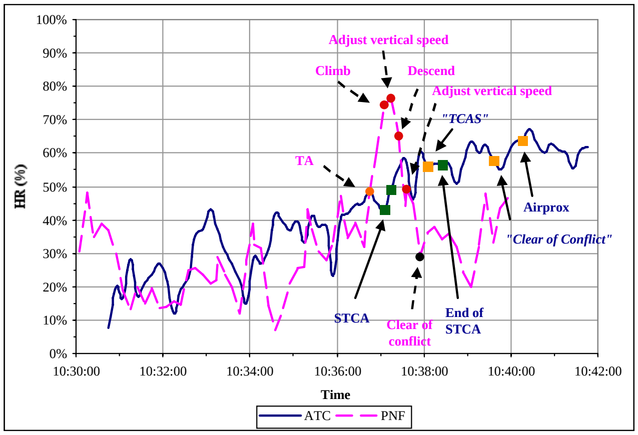


Figure 1. *HR expressed as a percentage of the reference (1st percentile of the total recording) during a TCAS sequence for one pilot (PNF) and one ATC*

Then, HR progressively decreased even with the two subsequent RA suggesting an adaptation of the physiological stress to the situation. The level returns to the initial level (around 30%) after the "Clear of Conflict" announce. For the ATCO, HR increased progressively after the STCA and reached a maximum (>60%) after the TCAS and airprox reporting by the aircrew. In most of the simulations, a similar pattern was observed with some variability in the magnitude of variations. This result confirms that, even in a part task simulator, the scenarios and the environment are able to induce a significant stress effect. In some cases, stress induced changes in behaviour. In one simulation, after the aircrew had solved a multiple RA sequence, a second TA appeared while the crew was resuming normal navigation. This TA was not detected by the aircrew, and even during the auto-confrontation they had difficulties to recognize this event. This suggests a "post-stress" or a "slacking" effect that reduced the available resources of the crew. A systematic analysis is being carried out on the relationship between physiological manifestations of stress and some behavioural changes that occurred during the simulations.

Thematic Analysis

From the data collected, two topics have been selected as relevant from a Human Factors and operational point of view:

- situational awareness,
- aircrew-ATCO cooperation and communications.

Situational Awareness (SA)

SA has been analysed regarding four main issues:

- data collection,
- timing of the TCAS sequence,
- control over the situation,
- common perception of conflicts by aircrew and ATC

Data collection. Since its introduction, TCAS has introduced a major change in the perception of traffic situation by aircrew. In fact, surrounding traffics are continuously displayed on the ND (CDTI). Therefore, aircrews now try to build an overall picture of the traffic situation based on this information while in the past this was only done through the hearing of the ATC communications (party line). This may impact the R/T communications, even before the TCAS issues an alert. The following examples of aircrew messages to the ATC during the simulations were recorded before and during TA's (most are translated from French):

- Before a TA : *"we've got a traffic", "we've got an aircraft", "traffic TCAS", "you've got a traffic information ?".*
- During a TA : *"we've got a TCAS", "TCAS alert", "we've got a visual", "we've got a visual TCAS" "we have it on TCAS"*

These messages were intended to ask for traffic information or were an answer to an ATC clearance or a traffic information given by the ATC. They are not covered by any procedure or rule and may interfere with the ATC work and induce misunderstanding. For example, the word "visual" may be understood by the ATC as "I have a visual contact on the traffic" or "Traffic TCAS" can be understood as "we've got a RA". The display of traffic on the ND may also lead to false interpretation. For example in the Orly scenario all pilots have seen the traffic as the aircraft ahead on the approach, which was not the case. This misinterpretation has a direct impact on aircrew SA and may lead to incorrect maneuver in case of RA (as it happened in the real situation).

The timing of the TCAS sequences. The analysis of TCAS sequences reveals a large variability in the timing of the TCAS events. In this study, the duration of TA varies from 2 sec to 38 sec. In one case, a RA occurred without being preceded by a TA. The collective debriefing showed that most participants are not aware of this large variability. The absence of TA leads to a situation where the aircrews could not be properly prepared to respond to the RA. In this case, the procedure which is normally followed after a TA in most airlines (the captain announcing "*I (or you) have the control*", switching off the Flight Director) cannot be applied. The high unpredictability of the TCAS sequence impacts SA as prevision and anticipation play a major role in the building process of SA (Endsley, 1998).

The control over the situation by the aircrew. After each scenario, the participants were asked to rate how difficult it was to evaluate the situation and whether they felt they started to loose the control over the situation. Table 1 shows the results of these questions.

		No	Yes
Did you find difficult to assess the situation ?	PF	11	9
	PNF	16	4
Did you felt that you were losing control of the situation ?	PF	20	0
	PNF	20	0

Table 1. *Evaluation and control of the situation by the aircrew*

Results show that the feeling of a global situation assessment is higher among the PNF than for the PF. This can be explained by the fact that PF are mostly focussed on the active following of the RA and are not seeking to have an understanding of the situation. The following statements of PF's during the debriefings confirm this attitude: "You cannot react according to what you understand"," I don't know what happened" "I do not remember to descend", "I focused on the IVSI [NB : where the RA is displayed on Airbus aircraft], I do not look at the ND". During the debriefing, most of pilots stated that the RA TCAS are too unpredictable and that it is preferable to concentrate on the execution of the manoeuvre. In this context, they do not expect or seek traffic information from the ATC.

Common perception of conflicts by aircrew and ATC. One of the most striking results from the collective debriefing was the large shift in the perceptions of ATCOs and aircrews on the same situations. The auto-confrontation of the participants with the video recordings showed that most ATCOs are not aware of how the TCAS is displayed in the cockpit. Aircrews are also not informed about the ATC tools, especially regarding the functioning of the STCA and the characteristics of radar display (precision and refreshment rate). This was confirmed by the results of 2 questions asked to the aircrews and ATCOs (Table 2). These questions have been asked only for the Biarritz scenarios where ATCOs were not aware of the aim of study and did not expect the situation at all.

The most striking results are the large number of negative answers (11 out 20) and the uncertainty of the PNF (4 answers "don't know" out of 5). This shift is mainly due to the different and independent tools that are used by ATCOs and aircrews, e.g. time shift between STCA and TCAS. This leads to a lack of common perception of the situation which may interfere in the communication and cooperation between ATCOs and aircrews in these demanding situations.

		No	Yes	Don't know
To the aircrew: *Do you think you had a common perception with ATCO?*	PF	1	0	4
	PNF	5	0	0
To the ATCO: *Do you think you had a common perception with aircrew?*	ATC1	3	1	0
	ATC2	2	0	3

Table 2. *Feeling of a common representation by ATC an aircrew*

The Aircrew-ATCO Communications

In this section, the main results regarding the communications between ATCOs and aircrews are reported. The results are presented both for the messages from aircrew to ATCO and from ATCO to aircrew.

Aircrew notification The only way for the ATCO to be informed of a TCAS resolution is through the notification of the RA by the PNF. The airline procedure provides only 2 messages, whatever the RA issued: "TCAS climb" or "TCAS descend". In this study, for simple RA such as Climb or Descend, the observed messages are consistent with the procedure which is, in this case, clear and appropriate. For the other RA a large variability of phraseology is used, with sometimes some ambiguous. For example, some pilots used the message "TCAS descend" to report an Adjust Vertical Speed RA, although this RA always means a decrease of vertical speed that may occur while the aircraft is climbing. This raises the issue of the alert "Adjust vertical speed" which does not give directly the sense of the RA and, as a consequence, the way the pilot can report it to the ATC.

ATCO instructions Since the accident of Uberlingen (BFU, 2004) both aircrews and ATCOs are aware that aircrews must follow their RA and that ATC should not give any clearance to the aircrew. However, on the 5 scenarios that have been played where the ATCOs were involved, 2 ATC clearances have been given to aircrew who followed these clearances. In these two cases, the ATC clearance was given because the ATCO was trying to avoid a conflict with another aircraft. In one case, the clearance happened while the ATCO thought that the conflict is solved, for the other, the clearance was compatible with the RA TCAS. The critical aspect is that the initial RA could be followed by another RA which may be incompatible with the clearance.

Discussion and Conclusion

The results obtained in this study show that even a partial simulation of tasks was able to reproduce TCAS events, stress and behaviors that raise several human factors issues that could not be revealed in incident reporting. The simulation conditions enabled producing the temporal pressure and stress that is inherent in the TCAS sequences. The assessment method that was developed for this study, gathering physiological recordings, observations, verbalization and questionnaires showed its strength to detect and analyze the critical human factors issues to be addressed in the future. These issues have to be considered in the aircrew-ATCO relation and not only at one level. To summarize these issues, the TCAS sequences can be represented as a “parenthesis” in the normal aircrew-ATC communication and cooperation (figure 2). This figure depicts the several events and sequences that follow one another. The upper part represents the TCAS events occurring in the cockpit, the lower part the ATC side and in-between the air-ground communications. As it is shown, the effects of TCAS occur before the TA, when a traffic is displayed on the ND. This leads to a change in the communication with potential interferences and disruptive effects as it was reported earlier (Benhacène, 2001 ; Walsh, 1997). The subsequent sequence starts when a TA occurs. This period is critical for the aircrew as it is intended to prepare them for a potential RA. One of the main issue related to this period that was revealed by the study is the very large variability of the timing of the sequence: from very short (even in one case, with no TA) - which does not allow the crew to apply the expected procedure and be mentally prepared to react- to long periods where the preparation can diminish progressively. As airborne and ATC systems are independent, additional interferences can occur at this moment due to the STCA triggering which may induce actions from the ATCO. When the RA occurs, a critical period is starting (T1). As long as the aircrew has not reported the RA, the ATC has no means to be informed that the TCAS has issued an alert.

This creates a very sensitive situation where ATC may still give clearances that can be very disruptive for the aircrew. The reporting of the RA by the aircrew is expected to open the parenthesis in the aircrew-ATCO communications. However, as demonstrated by our results the reporting is sometimes inexistent, late or ambiguous. The “Clear of Conflict” (CoC) message from the TCAS starts another critical period (T2). As for T1, as long as it is not reported by the aircrew, the ATCO is ignorant of the end of the RA. This raises a transfer of liability issue between the aircrew and the ATCO: who is responsible for the separation of aircraft? The report of the CoC by the aircrew to the ATC closes the parenthesis, the aircrew normally returning to the initial clearance, and resuming normal navigation (auto-pilot ON, flight director ON). As it was shown in the results, these tasks and a potential slack in attention due to the stress experience during the RA may have potential impact in this period reducing the attention on subsequent TA.

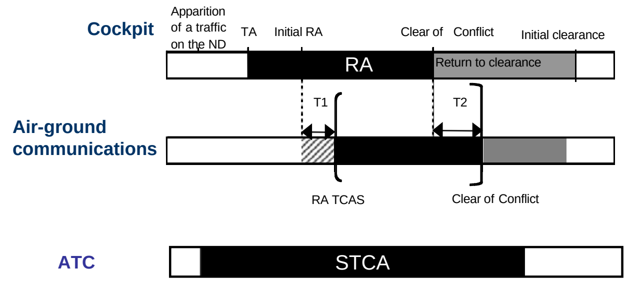


Figure 2. *The parenthesis in the aircrew-ATCO communications in the TCAS sequence*

Most of participants (pilots and ATCOs) stated that this type of simulation and common debriefing allowed them to better realize the operational issues and difficulties in these time-critical situations: some had a clear understanding of TCAS and associated procedures but no operational experience. They were surprised to have performed away from their understanding under time pressure and they noticed the consequences of their action on the other’s job (ATCO or aircrew). So this represents a step forward as far as training is concerned into practice for the training process. Further analyses of the data are currently conducted in order to get a systematic analysis of errors.

A second round of simulations was conducted in autumn 2004: some changes were applied to scenarios in order to keep the ATCOs in their operational role. This led to some new situations and opened some new issues about these very short intensive periods. From the whole results and discussions of both sessions, some solutions will be suggested, which may reinforce or question present studies related to TCAS improvement. One of the most encouraging outputs is the method that was used to tackle the human aspects of the air-ground integration and could be use for the evaluation of solutions such as the RA downlink (Broker, 2004): it is a valuable complement to other approaches that have already been conducted: incident analysis, simulations involving only one side (RADE1, 2004), or field evaluations (Walsh, 1997). It may also be a valuable complement to present training methods, which does not require outstanding technical means.

References

Benhacène, R. (2001). Avionics and their perception by Air Traffic Controllers. Proceedings of the Fourth International Air Traffic Management R&D Seminar ATM-2001, Santa Fe (New-Mexico, USA) Dec 3rd-7th.
Bundesstelle für Flugunfalluntersuchung German Federal Bureau of Aircraft Accidents Investigation. (2004). Investigation report AX001-1-2/02.

Brooker, P. Thinking about downlink of resolution advisories from airborne collision avoidance systems. Human Factors and Aerospace Safety 4(1), 49-65.

Cabon, P., Favresse, A., Mollard, R. (2003). Impact du TCAS sur l'activité des équipages et sur la relation équipages-contrôleurs. Constats et recommandations. Paris, France: Laboratoire d'Anthropologie Appliquée, Doc. AA 458/2003.

Endsley, M.R. (1988). A construct and its measurement: the functioning and evaluation of pilot situation awareness, NOR DOC 88-30, Hawthorne (Cal.), Northrop Corporation, 1988.

Hollnagel, E. (1998). Cognitive Reliability and Error Analysis Method, Elsevier, Amsterdam.

Powell, D., Baldwin, T. (2002). European ACAS operational monitoring – 2000 report. EEC report No. 373.

RADE1- RA Downlink Experiment – Final report, Eurocontrol, may 2004.

Walsh, J. (1997). TCAS RA Downlink Field Evaluation Results – AND-720 MITRE.

Wickens, C.D. (1992). Engineering psychology and human performance, New York, HarperCollins.

Acknowledgment

The authors would like to express their gratitude to the CENA technical team that made these simulations possible, especially to Christelle Pianetti, Remy Lafage and Jean-Paul Imbert. The scenarios were selected thanks to J-D. Blanchecotte, B.Giosca and Didier Gallard from the Service du Contrôle du Trafic Aérien. Finally, this study would not have been possible without the active participation and collaboration from the controllers and the aircrews

PERFORMANCE EVALUATION OF A COMPUTATIONAL MODEL OF EN ROUTE AIR TRAFFIC CONTROL

Todd J. Callantine
San Jose State University/NASA Ames Research Center
Moffett Field, CA

This paper describes a model of en route air traffic control and presents the results of a performance evaluation of computational air traffic controller agents based on the model. The purpose is to better understand the representations, heuristics, and processes that expert air traffic controllers use and develop agents useful for air traffic management concept development and safety/risk analysis. The results show the agents control low-to-medium traffic levels effectively. The research was supported by the NASA Aviation System Capacity Program and the FAA/NASA Aviation Safety Program.

Introduction

Today's air traffic management (ATM) system is highly safe and robust, but it cannot sustain current capacity limits, inefficiencies, and adverse environmental impacts over the long term. Researchers are therefore investigating new ATM concepts to address these problems. The complexity of the ATM system makes developing new concepts challenging. Researchers must address a broad range of issues—automation functionality and operator interaction, operational scenarios, and training. Simulations with computational agents offer an attractive complement to development through iterative human-in-the-loop simulations.

Several recent research efforts address air traffic controller models. For example, Niessen, Eyferth, and Bierwagen (1999) studied how experienced controllers assess traffic situations. Niessen and Eyferth (2001) then used a computational cognitive model based on the ACT-R framework to study how controllers construct a 'picture' of the traffic situation. Other research has investigated control strategies (Nunes and Mogford, 2003) and conflict detection and resolution rules (Mondoloni, 1998). Models have been developed to assess control techniques (Krozel, Peters, Bilimoria, Lee, and Mitchell, 2001), produce predictive performance measures (Leiden, 2000), and enable decision support (Hexmoor and Heng, 2000).

ATM safety and efficiency studies have also been conducted with computational cognitive models. For example, AirMIDAS has been used to analyze the safety of new alerting systems (Pritchett, Lee, Abkin, Gilgur, Bea, Corker, Verma, Jadhav, Reynolds, Vigeant-Langlois, and Gosling, 2002) and the effects of proposed changes to practitioner roles and responsibilities (Corker, Gore, Fleming, and Lane, 1999). Cognitive agent models of conflict resolution in distributed ATM have also been developed (Harper, Guarino, White, Hanson, Bilimoria, and Mulfinger, 2002).

This paper describes a model and its implementation as a computational agent that functions as a radar (R-side) controller controlling traffic in a single sector. The model approximates controller behavior using heuristic methods rather than optimization methods. The research aims to better understand the representations and processes air traffic controllers use and refine agents useful in advanced ATM concept development and safety/risk analysis. After describing the model and its implementation in en route controller agents, the paper describes a performance evaluation with three agents controlling arrival traffic in adjoining sectors. Additional detail is provided in Callantine (2002b).

Model and Computational Architecture

Figure 1 shows the information flows within an agent and its interactions with other agents and a traffic simulation via a 'simulation hub.' Agents issue clearances to simulated aircraft, initiate handoffs to other agents, and accept handoffs from other agents using messages passed through the simulation hub. Figure 2 shows a screen snapshot of an agent controlling traffic. The following sections describe the model components and processing.

Activity Model

A Crew Activity Tracking System (CATS) activity model serves as the basis for the air traffic controller agents (Callantine, 2001). The model represents the high-level structure of the air traffic control task. Each air traffic controller agent uses the CATS activity model shown in Figure 3. The model represents activities hierarchically, down to the action level, and includes conditions that specify when each activity should preferably be performed.

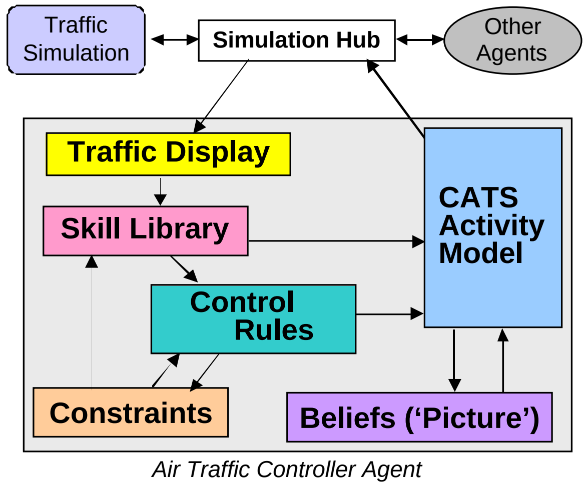


Figure 1: *Information flows within and between air traffic controller agents.*

The model in Figure 3 can be thought of in three parts. The first is the Maintain situation awareness activity, and its children, Monitor traffic display and Scan aircraft. These activities are devoted to gathering information from displayed traffic information. The second is the Determine aircraft to work activity, which represents selecting a traffic control problem to address from those currently identified. The third portion is a collection of Manage *X* activities that are performed based on the outcome of the Determine aircraft to work activity. Thus, the model is similar to conceptual air traffic controller models with situation assessment, planning, and execution modules (e.g. Davison and Hansman, 2003).

Agents exhibit a 'flow of activity' that hinges on the Determine aircraft to work activity. Executing this activity identifies the next aircraft (or 'cluster' of aircraft) that the agent should address according to a static set of priorities. In plans with multiple steps (e.g., vector an aircraft off its route, then to a route-intercept heading, then back on its flight plan route), later steps depend on earlier steps for their success. The highest priority is therefore to implement plans whose execution conditions are currently satisfied. Planning to solve conflicts is second, planning to solve spacing problems third, and issuing descent clearances fourth. Handoff acceptance and handoff initiation are the lowest priority. The priority

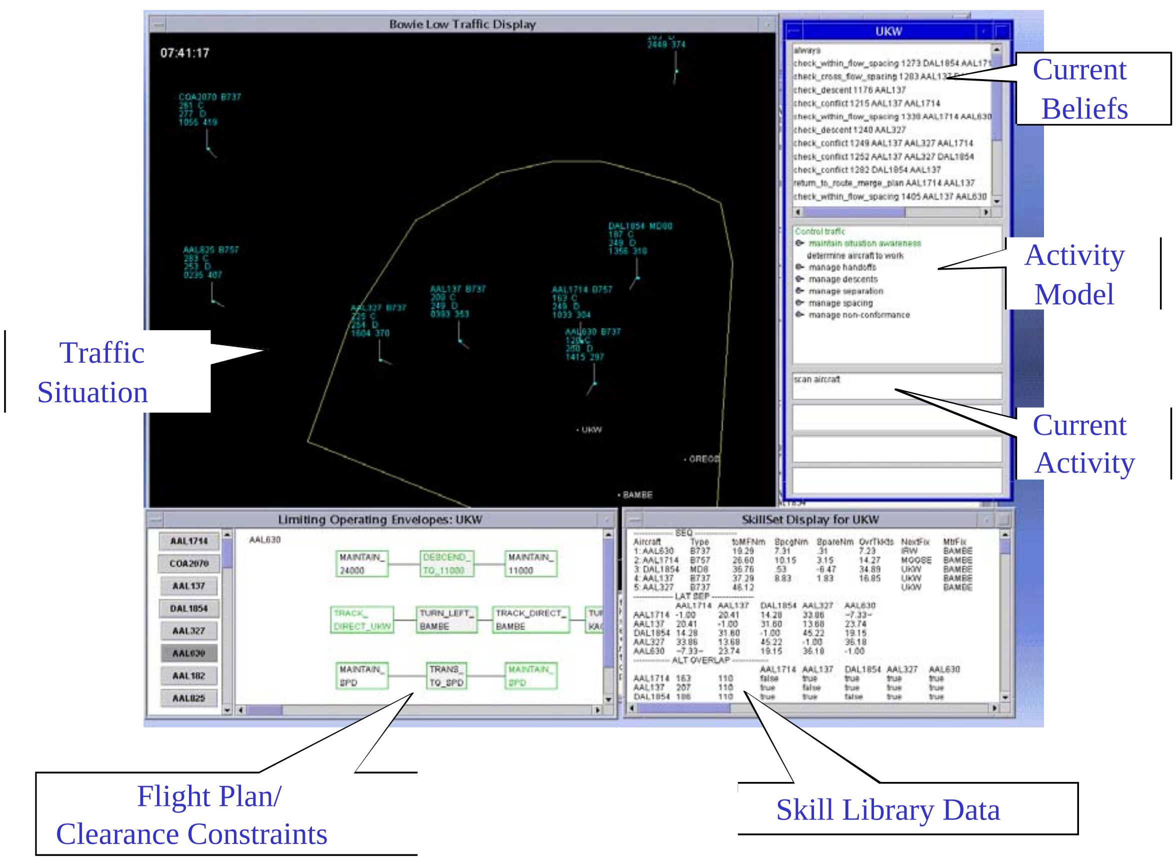


Figure 2: *Screen snapshot.*

- Maintain situation awareness
  - Monitor traffic display
  - Scan aircraft
- Determine aircraft to work
- Manage handoffs
  - Accept aircraft
    - Accept handoff
    - Roger check-in
  - Initiate handoff
    - Inform other controller
    - Issue frequency change
- Manage descents
  - Issue descent clearance
- Manage separation
  - Evaluate separation clearance options
  - Issue separation clearance
- Manage spacing
  - Evaluate spacing clearance options
  - Issue spacing clearance
- Manage nonconformance
  - Re-issue clearance

Figure 3: *CATS activity model.*

structure enables agents to reasonably approximate 'chunking' of air traffic controller behavior. As an example, controllers are sometimes observed to issue several clearances to separate a cluster of aircraft, then accept several handoffs in succession.

Beliefs

The agents maintain beliefs about the current task context and current traffic situation. Agents transform their belief set by performing activities, in accordance with the theory that all salient operator activities in complex human-machine systems involve transforming or communicating contextual information. For example, performing a perceptual activity entails transforming information found in a representation of the appropriate visual or auditory 'display' into a set of beliefs about the information. Performing a cognitive activity entails modifying the agent's belief set to produce beliefs at different levels of abstraction, or beliefs that encapsulate the results of a decision making process.

Task context beliefs on the left side of Figure 4 appear in the conditions for performing activities in the CATS activity model. Depending on various traffic assessments, the agents add or remove different beliefs from their belief set. The last several beliefs ('know which…' and '…identified') correspond to the type of control problem identified in Determine aircraft to work. For example, if the Determine aircraft to work activity finds a conflict is the highest priority problem, an agent adds 'factors identified' to its task context belief set, which causes the agent to execute Evaluate separation clearance options on the next processing cycle. Executing this activity references the 'control rules' heuristics and results in a 'know which aircraft to clear' belief, which then triggers the Issue separation clearance activity.

The right side of Figure 4 lists beliefs about the current control situation, including memory for when problems were last addressed, and prospective memory for plans. By planning to issue a clearance to solve the conflict, rather than issuing the clearance right away, the agent has the option to adapt the plan or abandon it altogether if its execution conditions happen not to materialize. Retrospective memory about when problems were last addressed is also important because it takes time for traffic to reflect the effects of clearances. The 'check…' beliefs tell the agent to move on to lower priority problems until after the indicated time (Figure 4). Situation beliefs

Task context

Always
Display needs scanning
Looked at traffic display
Have aircraft to work
Know which aircraft to accept
Know which aircraft to hand off
Know which aircraft to descend
Factors identified (refers to conflict aircraft)
Spacing aircraft identified
Know which aircraft to clear (separate)
Know which aircraft to space
Know which aircraft is not conforming

Situation context

- *beliefs about current situation*
- *memory for 'problem status'*
- *prospective memory for plans*

Check_cross_flow_spacing [time] [aircraft]
Check_within_flow_spacing [time] [aircraft]
Check_conflict [time] [aircraft]
Check_descent [time] [aircraft]
Cross_flow_spacing [aircraft clusters]
Within_flow_spacing [aircraft clusters]
Conflicts [aircraft clusters]
Sector_aircraft [aircraft]
Plan_exec [aircraft]

Figure 4: *Task and situation beliefs.*

refer to individual aircraft or clusters of aircraft, according to the level of structural abstraction required (Davison and Hansman, 2003).

Agents also maintain other important information via encoded Java™ objects and variables. The most important of these are 'role bindings' for aircraft, which provide a general way to specify a frame of reference for the application of heuristics. When agents execute the 'monitor traffic display' activity, they apply encoded skills to 'bind' aircraft to roles (e.g., front, frontSequence, etc.). For each bound role, the agents also access skills that assign a bit-vector of fuzzy-valued attributes (e.g., tooClose, atSameAltitude, etc.).

Control Rules and Plans

A collection of heuristics determines the control techniques to use to achieve proper spacing or separation (Figure 5). Spacing heuristics relate to establishing a desired in-trail distance, while separation heuristics resolve conflicts. In this research, spacing problems can by definition be solved using speed clearances, while separation problems by definition require heading vectors. Separation heuristics are differentiated according to whether aircraft are merging or not. The control rules use role bindings to reference other aircraft.

Planning is crucial for solving separation and spacing problems. The heuristics address the aircraft currently bound to roles; however, other aircraft may also be in conflict. Allowing the agents to develop plans for all conflicting aircraft before issuing any clearances means agents first execute plans whose execution conditions are met first. Figure 6 shows plan steps in each dimension (grayed-out plan steps were replaced with immediate clearances in the evaluation study). Figure 6 also shows examples of plan-adaptation conditions for lateral plans. Each plan contains roles (e.g., front, etc.) bound to the plan and a 'planned time' for executing the plan. Plans may simply be

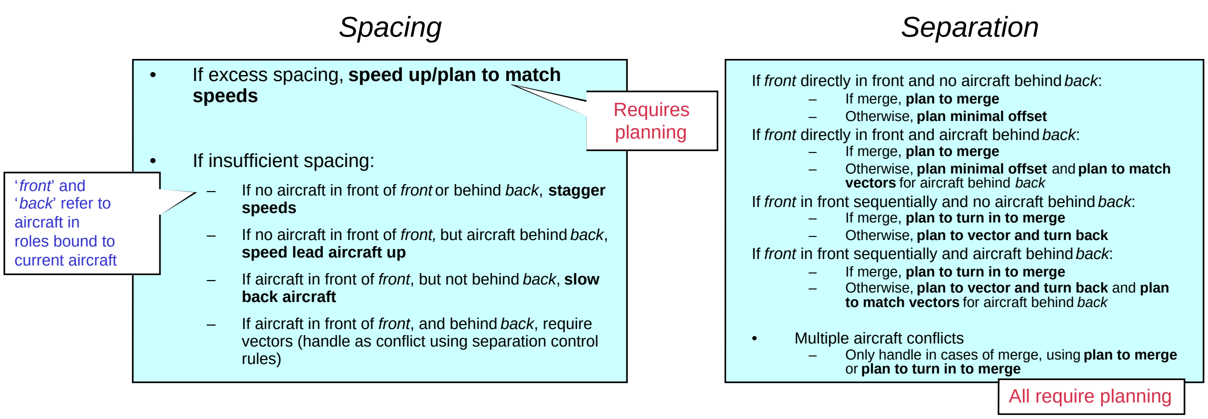


Figure 5: *Spacing and separation control rules.*

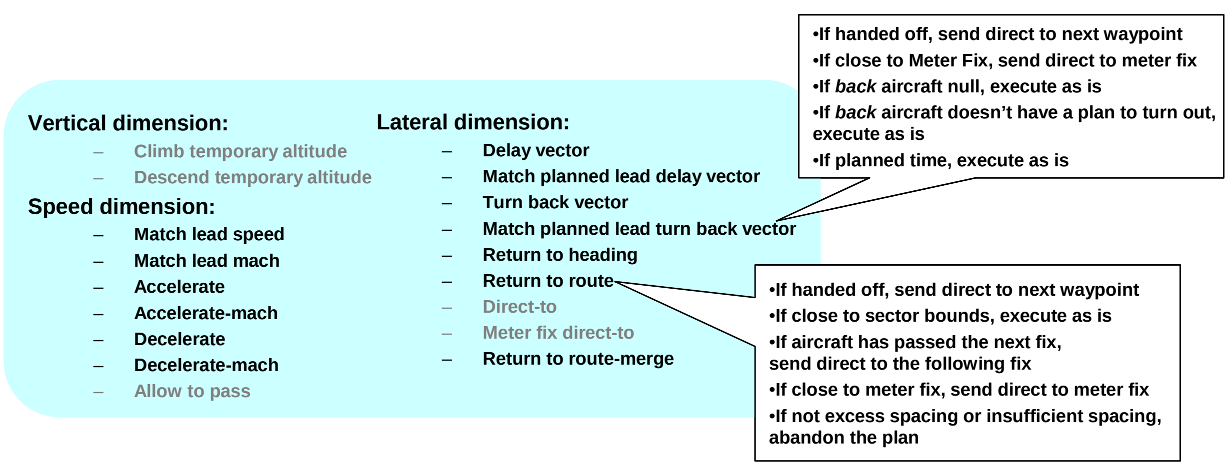


Figure 6: *Plan steps and examples of adaptation/execution conditions.*

executed at their planned time if no adaptation conditions are met.

Skill Library

The 'skill library' is collection of encoded methods that enable agents to perform low-level pattern recognition and display-based decision-making. Examples include determining the lead aircraft for an aircraft of interest, determining the precise heading to issue when a heading clearance is called for, or assessing the distance between two aircraft. Skills figure prominently in determining which aircraft to work, applying control rules, and monitoring plan adaptation/execution conditions.

Constraints

Each agent maintains a representation of operational constraints on each aircraft (see Figure 2) in its 'area of regard'. Constraints derive from the aircraft's flight plan and amendments to it specified by clearances (Callantine, 2002a). The constraint representation enables agents to monitor conformance with clearances and predict future behavior (e.g., time remaining until an aircraft should maneuver).

Traffic Display

The traffic display is a representation of the information available on a controller's scope (see Figure 2). Skills operate on the traffic display information to assess the traffic (see Figure 1).

Method

A performance evaluation was conducted with three agents controlling simulated arrival traffic in en route airspace in real time. The evaluation compared number of loss-of-separation events (less than 5 nm of lateral separation and less than 1000 ft vertical separation) with and without full agent control.

Airspace

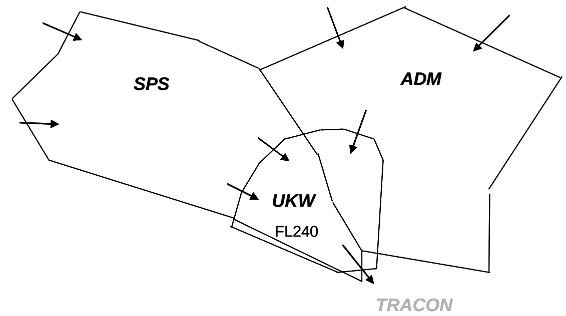


Figure 7: *Airspace and arrival traffic flows.*

Two agents simultaneously controlled traffic in high altitude sectors SPS and ADM; another agent was responsible for merging the arrival flows in the low altitude sector UKW (Figure 7).

Traffic Scenarios

Nine scenarios were adapted from scenarios that were being used in other NASA ATM research. The scenarios represented a range of traffic conditions. Each of the nine scenarios was run first in a 'no control' condition with agents only issuing descent clearances, so that aircraft simply arrived on their nominal flight plan arrival trajectory. Each scenario was then run again with the agents issuing clearances.

Results

Figure 8 summarizes the performance evaluation results. The agents handle spacing problems in the high altitude sectors (SPS and ADM) well. The agents are less adept at handling merge problems in UKW. More loss-of-separation events occurred in dense-traffic scenarios with poorly conditioned arrival flows (scenarios 7-9). In no case did the agents produce more loss-of-separation events than the uncontrolled (descent clearance only) condition.

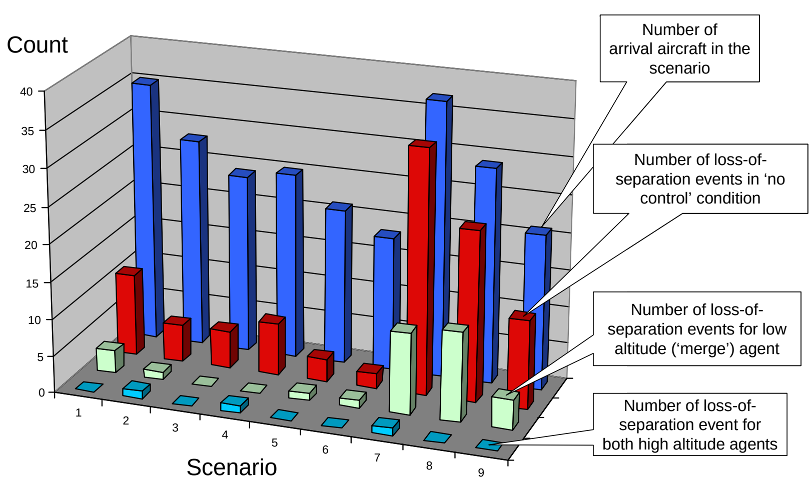


Figure 8. *Scenario traffic counts and loss-of-separation events.*

Conclusion

The agents performed reasonably well considering the difficulty of the air traffic control task. The knowledge representations and processing scheme embodied in the agents are elicited from observations and anecdotal evidence about how human controllers operate. The control rules, plans, adaptation/ execution conditions, and prioritization of control problems therefore may not be appropriate in every situation. Because the study did not include professional human air traffic controllers, suitable validation measures are not available. In addition to validated control knowledge, the results suggest that better predictions and intentional focus would improve the 'picture,' and in turn, overall agent performance.

Current research is addressing enhancements to the air traffic controller model and computational architecture. The enhanced agents are designed to control traffic in terminal radar approach control rather than en route airspace. Human controller performance data is available for the same traffic scenarios to be used for agent testing, which will enable detailed validation studies.

Acknowledgements

This research was supported by the NASA Aviation System Capacity Program and the FAA/NASA Aviation Safety Program. Thanks to Everett Palmer, Thomas Prevôt, and Paul Lee for providing useful insights.

References

Callantine, T. (2001). Agents for analysis and design of complex systems. *Proceedings of the 2001 International Conference on Systems, Man, and Cybernetics*, 567-573.

Callantine, T. (2002a). A representation of air traffic control clearance constraints for intelligent agents. *Proceedings of the 2002 IEEE International Conference on Systems, Man, and Cybernetics,* #WA1C2 (CD-ROM).

Callantine, T. (2002b) CATS-based air traffic controller agents. NASA Contractor Report 2002-211856. Moffett Field, CA: NASA Ames Research Center.

Corker, K., Gore, B., Fleming, K., and Lane, J. (2000). Free flight and context of control: Experiments and modeling to determine the impact of distributed air-ground air traffic management on safety and procedures. *Proceedings of the 3rd USA/Europe Air Traffic Management Research and Development Seminar*, Naples, Italy.

Davison, H., and Hansman, J. (2003). Using structure as the basis for abstraction in air traffic control. *Proceedings of the 12th International Symposium on Aviation Psycholog*y. Columbus, OH: The Ohio State University.

Harper, K., Guarino, S., White, A., Hanson, M., Bilimoria, K., and Mulfinger, D. (2002). An agent-based approach to aircraft conflict resolution with constraints. AIAA Paper 2002-4552, Reston, VA: American Institute of Aeronautics and Astronautics.

Hexmoor, H. and Heng, T. (2000). Air traffic control and alert agent. *Proceedings of the 4th International Conference on Autonomous Agents*, Berlin: Springer-Verlag, 237-238.

Krozel, J., Peters, M., Bilimoria, K., Lee, C. and Mitchell, J. (2001). System performance characteristics of centralized and decentralized air traffic separation strategies, *Proceedings of the 4th USA/Europe Air Traffic Management Research and Development Seminar*, Santa Fe, NM.

Leiden, K. Human performance modeling of en route controllers (RTO-55 Final Report). Boulder, CO: Micro Analysis and Design, Inc., 2000.

Mondoloni, S. (1998). Development of an enroute conflict resolution rulebase for the reorganized air traffic control mathematical simulator. NARIM-A09008-01, Washington, D.C.: CSSI, Inc.

Niessen, C. and Eyferth, K. A model of the air traffic controller´s situation awareness. *Safety Science, 37* (2001), 187-202.

Niessen, C., Eyferth, K. and Bierwagen, T. (1999). Modelling cognitive processes of experienced air traffic controllers. *Ergonomics, 42*(11), 1507-1520.

Nunes, A., and Mogford, R. (2003). Identifying control strategies that support the 'picture', *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*, 71-75.

Pritchett, A., Lee, S., Abkin, M., Gilgur, A., Bea, R., Corker, K., Verma, S., Jadhav, A., Reynolds, T., Vigeant-Langlois, L., Gosling, G. (2002). Examining air transportation safety issues through agent-based simulation incorporating human performance models, *Proceedings of the 21st Digital Avionics Systems Conference*, 7.A.5-1-7.A.5-13 (CD-ROM).

DEVELOPING PILOT SKILLS IN CRM ASSESSMENT: AN EFFECTIVE TOOL FOR THE CHECK PILOT

Jose Manuel Ponz Cantó
Jefe Unidad CRM y Factores Humanos Iberia
(Head of Human Factors and CRM Iberia Airlines)

The European Joint Aviation Authorities eventually opted for making the check pilot responsible for CRM skills assessment. Along the way, several options were considered; NOTECHS, a version of LOSA, or even special tests for CRM skill demonstration.

Eventually, it was decided that CRM skills should not be evaluated outside of the pilot´s flying environment (we believe quite rightly), and such skills, or behavioural markers should only be assessed during flight checks.

Thus, the problem was apparently solved. However, no provisions were made for the training of check pilots in CRM assessment, and no special training other than required in the ATP license syllabus was devised.

Therefore we found that pilots were being assessed in CRM skills by other pilots with no specific training in CRM, and more often than not, with little training in human factors at all, due to the generally high experience, old timer status of check pilots, who therefore obtained their licenses at a time when no provision or hardly any was made for the teaching of human factors.

In view of this situation, the CRM and human factors department at Iberia have designed, under my supervision and guidance, a program that specifically deals with the issue of CRM skills assessment training for instructors and check pilots. It is a highly condensed, very practical and participative syllabus of only 8 hours, in which we first demonstrate the difficulty of assessing interpersonal skills, then the disparity of evaluation criteria, and then show some techniques and procedures for accurate and standarized (as much as it can posssibly be standarized) assessment. Finally we ask the students to assess a particular case (videotaped or occassionally role played) and we compare the distribution of scores with the one they did at the beginning of the class, thereby constantly evaluating our own effectiveness in teaching .

Indeed, we would like this syllabus to be of longer duration; however, due to the commercial pressure on Airlines, and given the lack of legislation requiring this particular training, we are quite satisfied with the adoption by Iberia of this measure , which guarantees check pilot´s and instructors´ awareness of the processes involved in human factors assessment.

DEVELOPMENT AND INTEGRATION OF HUMAN-CENTERED CONFLICT DETECTION AND RESOLUTION TOOLS FOR AIRBORNE AUTONOMOUS OPERATIONS

Riva Canton
QSS Group Inc.
Moffett Field, CA

Mohammad Refai
UARC
Moffett Field, CA

Walter W. Johnson and Vernol Battiste
NASA Ames Research Center
Moffett Field, CA

Today's crowded airspace burdens both the pilot and controller with a heavy workload pertaining to the maintenance of conflict-free flight. Conflict detection and resolution (CD&R) tools have become a key element in modern flight systems and future airspace concept simulations. In this paper we describe an automated resolution tool that was developed at NASA Ames Research Center as part of an experimental evaluation of the Distributed Air-Ground concept. The tool is based on an analysis of conflict geometry and was developed as an intent (i.e. flight plan) resolution system. A key simplifying concept used in the development of airborne automated resolutions is the notion of "Rules of the Road" - a set of rules that uniquely assigns responsibility for the mitigation of a conflict. This paper outlines the challenges in developing such an automated resolution tool, as well as the lessons learned and the limitations observed.

Introduction

Free flight allows aircraft greater flexibility in en route maneuvers but shifts the responsibility for maintaining safe separation with other aircraft onto the pilot. With the shift in responsibilities, a flight deck tool is required in order to aid the flight crew with the tasks of maintaining separation. This tool should detect conflicts far in advance so that pilots can respond to conflict alerts in a strategic manner. This approach is in contrast with the reactive, tactical response elicited by the current Traffic Collision Avoidance System (TCAS), whose alerts are short range and immediate. To study this concept, the Flight Deck Display Research Laboratory at the NASA Ames Research Center has developed a Cockpit Display of Traffic Information (CDTI) system that is integrated with a Conflict Detection and Resolution (CD&R) tool. Based on flight path "Intent", the CD&R tool detects conflicts up to 12 minutes in advance and automates conflict resolutions by presenting to the pilot a list of pre-computed maneuvers that will result in a "de-conflict" prior to the time of loss of separation (LOS). In June of 2004, as part of the Distributed Air-Ground Traffic Management (DAG-TM) research program, research teams at the NASA Ames Research Center and Langley Research Center conducted a joint experiment to investigate the operational feasibility of the En Route Free Maneuvering concept, also known as Concept Element 5 (CE 5). Central to the CE 5 study was the idea of increasing airspace throughput by shifting more responsibilities to the airborne systems for maintaining separation. In particular, aircraft equipped with CD&R tools and flying autonomously are responsible for maintaining separation from other autonomous aircraft and from aircraft that are under Air Traffic Control (ATC) management ("managed" aircraft). The sections below discuss the implementation of the CD&R tool, experimental trials, evaluation, and future research and development in this area.

Implementation

Design Goals

The overall objective is to cultivate a flight deck system that will promote the efficacy of free flight. The effectiveness of CD&R tool from a human-factor perspective can be studied using a laboratory prototype of the system. Long term issues involving CD&R tool-design for the next generation flight decks can also be addressed. The primary design goal is that it must be human-centered, and an extension of a pilot's decision faculty. It should require no attention from the pilot in the absence of a conflict alert and it should not inundate the pilot with complex resolution activities when conflicts are detected. This system will serve as a strategic planner that provides the pilot with greater degree of freedom in terms of time and maneuver-options when confronted with conflicts. A near instantaneous response to a user action is crucial to the effectiveness of a CD&R tool. Therefore, system performance is a major consideration.

Conflict Detection Algorithm

The conflict detection algorithm in the CD&R tool is an adaptation of the methods described by Yang and Kuchar (1997, 1998). The algorithm uses aircraft intent information to propagate current states forward in time. These projected flight trajectories are then used to search for conflicts with the ownship (the observer's aircraft hosting a CD&R tool). A conflict is defined as an incident in which the ownship's protected zone is penetrated by another aircraft (intruder). The protected zone is a cylindrical volume of space 5 nm in radius and 2000 feet in height. With the ownship at the center, the protected zone is projected out along its trajectory while searching for conflicts with other aircraft.

The core of the algorithm is built based on a probabilistic model, but it can be configured to become a deterministic model at run-time by reducing the sampling rate to N=1.

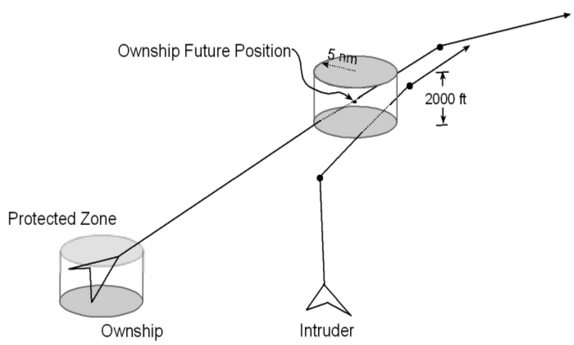


Figure 1. *A conflict occurs when intruder aircraft penetrates the ownship's protected zone.*

In the probabilistic approach, Gaussian and non-Gaussian distribution errors are introduced into the position, speed, and heading components of the aircraft states to model trajectory uncertainties. A Monte Carlo method is used to simulate the perturbed trajectories over N iterations. The probability of a conflict is the number of detected intrusions (or "hits") divided by N.

With upwards of 300 aircraft to process in the simulated airspace, performance is a primary consideration. Performance issues are mitigated in various ways. A number of filters are applied in order to screen out unlikely conflict candidates early in the process. Load management is accomplished through configurable sampling rate. A sampling rate of one second with 500 Monte Carlo iterations has been found to provide satisfactory results when combined with sample filtering. Using a 3.2 GHz dual processor and high speed graphic card at each simulation station, the system CPU budget is 25% for CD&R while graphical computation and other processes take up another 40%. Finally, the CD&R system is a standalone multi-threaded component; it can be deployed independently on a separate computer system to increase processing speed.

It should be noted that the solutions (computed conflicts) must be invariant. Specifically, a conflicting aircraft pair should see the same alert attributes (situational Awareness (SA) level, time to lost of separation (LOS), etc) from both sides.

Alert System and Symbology

Alerts are presented to the pilot through an escalating progression of alert conditions instead of an all-or-nothing approach, as would be the case for a TCAS resolution advisory. Alerts are categorized into three SA levels, with SA3 being the highest urgency and loss of separation imminence, and SA1 the lowest. In the probabilistic approach, an SA level is assigned by weighting the probability of a conflict with the corresponding Time Remained Prior to Loss of Separation (TLOS). The result is a mapping table shown in Figure 2.

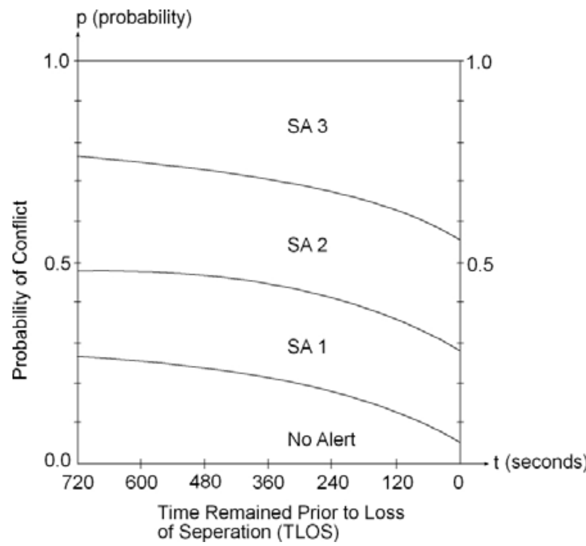


Figure 2. *Assigning SA levels in a probabilistic model - Mapping of probability against TLOS.*

Since uncertainty increases with time and distance in a predictor system, the probability of a conflict is therefore inversely related to the distance and time to the point of LOS. It is precisely this characteristic that facilitates a multi-leveled alert system. The probability of conflict becomes higher as aircraft approach LOS. Inspection of Figure 2 shows that an SA1 alert indicates medium probability with long TLOS to low probability with short TLOS; an SA2 alert indicates moderately high probability with long TLOS to medium probability with short TLOS; SA3 alert indicates high probability in general.

In the deterministic approach (N=1), no uncertainties are introduced. The multi-leveled transition depends on TLOS alone; staged at twelve minutes, eight minutes, and four minutes for levels SA1, SA2, and SA3 respectively.

Alert presentation to the crew employs various visual and auditory cues. At SA1, the ownship's symbol (default color is magenta) and the intruder's symbol (default color can be blue, green, or white) both turn to amber on the CDTI. If the intruder aircraft is out of display range, the "Alert" button "lights up" in yellow to alert the pilot that the intruder aircraft is not in view. By clicking on the "Alert" button, CDTI automatically zooms out to a larger range that brings the intruder aircraft into view. When the alert level escalates to SA2, an amber halo is superimposed on the conflicting-aircraft symbols. At SA3, an amber predictor pulse is projected along the flight paths, and an amber protected-zone-ring is projected out to the LOS position. Also at SA3, an audible chime is sounded. This transition from a subtle visual stimulus to a more salient one coupled with an audible sound is designed to cue the pilot as to the degree of urgency, thereby prompting the pilot to prioritize tasks.

Figure 3. *CDTI showing an SA3 alert level. Alert button lights up in yellow (bottom, second from left).*

When SA1 first appears at roughly twelve minutes prior to LOS, alert presentation cues the pilot that there is ample time to act and more options are available if action is taken immediately. When the alert level escalates to SA2 at roughly eight minutes to LOS, the pilot is reminded that there is a moderately high probability that a loss of separation is going to occur, and that the situation should be resolved within four minutes. When the alert level escalates to SA3 at roughly four minutes to LOS, a loss of separation is imminent - something has to be done immediately. Figure 3 depicts an SA3 alert level in the CDTI.

Concept of Conflict Probes

A probe is defined as a deliberate search for conflicts along an "Intent" trajectory. The primary "Current Probe" probes the current intended route and is active at all times. However, the CD&R tool has two additional probes: the "RAT (Route Analysis Tool) Probe" and the "Vector Probe". A dedicated Monte Carlo simulation powers each probe. The RAT is an independent component of the CDTI that provides a graphical user interface for modifying a flight path by inserting, deleting, and moving waypoints and leg segments of the existing flight plan. A detailed presentation of the RAT is beyond the scope of this paper. It will suffice here to characterize the RAT as a strategic planner for route modifications. When a modified route is proposed using the RAT (RAT route), a new probe is set off to search for conflicts along the proposed flight plan, thereby providing a level of confidence that the route is conflict free before committing to it.

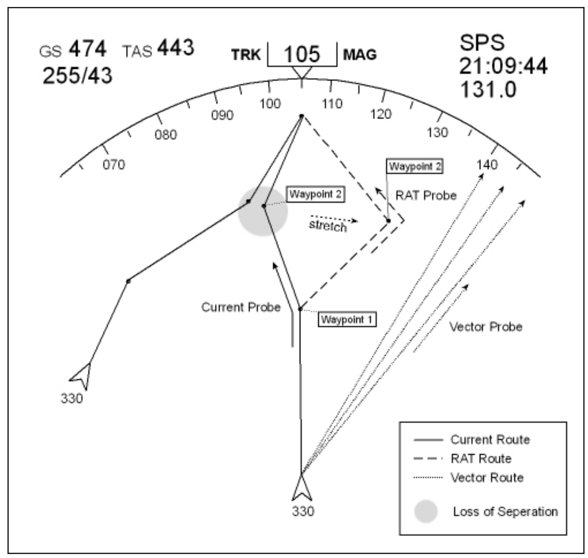


Figure 4. *Current, RAT, and Vector Probes along their respective routes on the CDTI display.*

The Vector Probe allows a pilot to probe for conflicts along an arbitrary heading. Dialing the heading on the Mode Control Panel activates the Vector Probe. A probe is set off to search for conflicts along the heading line as the pilot sweeps it across the display. This probe boosts the effectiveness of the CD&R tool, allowing it to support a free-flight environment in the truest sense. Figure 4 depicts the three conflict probes on the display.

Resolution Algorithm and Automation Display

The conflict resolution algorithm is an adaptation of the geometric optimization method presented by Bilimoria (2000). Efficient conflict resolution commands are computed for four different types of maneuvers: altitude change, speed change, heading change, and a combination of heading and speed change; these resolutions are presented to the pilot as proposed flight plans. For each maneuver type, two solutions with the least deviations from the nominal trajectory are selected. A maximum of eight solutions are provided when available. The computed resolutions are prioritized by their efficiency. As shown in Figure 5, a list of computed resolutions pops up when the "Res" button is clicked. The most efficient maneuver (least perturbation to the current trajectory) appears at the top of the list. The appropriate proposed flight plan is loaded into the RAT when the pilot clicks on one of the resolution options; this affords the pilot the opportunity to inspect and revise the selected resolution at will.

Figure 5. *Display of automated conflict resolutions (enlargement shows 3 maneuver options).*

In some cases, not all eight resolutions are available due to constraints such as TLOS, altitude restrictions, proximity of other aircraft, and FMS equipage (an FMS may not be able to implement a combined speed-heading maneuver, for example). The list of automated resolutions is dynamic. If no action is taken while the LOS point is approaching, these resolution options will expire one by one as they become invalid. The pilots can choose to ignore the automated resolutions and manually devise their own avoidance maneuvers.

The Rules-Of-The-Road Component (ROR)

The outcome of conflict detection is expected to be invariant and symmetrical between aircraft. Specifically, conflicting aircraft pairs should receive identical alerts if they have deployed the same CD&R tool. This can potentially lead to a race condition when both aircraft execute avoidance maneuvers concurrently, which if uncoordinated, may result in further conflicts that could become un-resolvable. To mitigate such situation, the CD&R tool incorporated "rules of the road" - a set of rules designed for coordinating collision avoidance in VFR flight.

ROR is a component of the CD&R system that automates the application of rules to a conflict situation. ROR relieves the flight crew from the distraction of having to mentally analyze the situation and apply the proper rule to arrive at a right-of-way conclusion. The right-of-way issue is settled by means of *burdening settlement*. In other words, ROR analysis identifies which aircraft has the burden of resolving a particular conflict.

When a conflict is detected, ROR analyzes the flight plans and the flight states of the conflicting aircraft at the point of LOS. A set of hierarchically ordered rules is then applied sequentially. A rule is found applicable only if the following complemental condition is satisfied: one aircraft must be non-compliant while the other is compliant with respect to that rule. If a rule is found to be inapplicable, then the next rule is applied and so on until the complemental condition is satisfied. The non-compliant aircraft is said to be the burdened aircraft and will be responsible for making trajectory modifications in order to resolve the conflict. The outcome of ROR analysis is a *burdening settlement* advisory that is issued to the two aircraft. Each settlement is accompanied by a short phrase (reason) that cites the particular rule leading to the settlement. By this automation process, only one aircraft is required to take action to resolve a conflict, thereby mitigating the potential danger of a race condition early on. Figure 6 shows multiple *burdening settlements* issued by ROR during multiple conflicts.

To avoid ambiguities induced by highly articulated flight paths, the ROR rules are applied at the point of LOS. The following is the list of hierarchical rules implemented in the ROR (definitions of these rules as well as an in-depth treatment on ROR are presented by Johnson, Canton, Battiste, and Johnson. 2005):

- IFR/AFR rule
- Altitude Rule
- Vectored Rule
- Left/Right Rule
- Level Flight Rule
- Descend/Climb Rule
- Overtake Rule

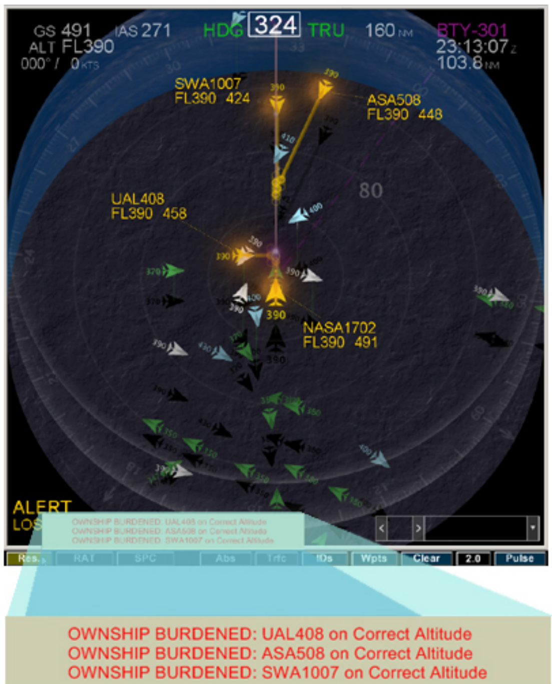


Figure 6. *Display of burdening settlements during multiple conflicts (enlargement shows reasons).*

Flight Deck Integration

To emulate full flight deck functionality on different platforms for the CE5 study, the conflict detection-capable CDTI was integrated into the Advanced Concepts Flight Simulator (ACFS) as well as the Multi-Aircraft Control System (MACS). The ACFS is a 6-degree-of-freedom full mission B737 flight simulator in the Crew Vehicle Systems Research Facility (CVSRF) at the NASA Ames Research Center. The MACS is a desktop computer flight simulation program that emulates the B777 flight deck controls. It was developed by the Airspace Operation Laboratory (AOL) at Ames (Prevot, 2002).

Experimental Trials and Evaluation

Conflict Detection and Alerting

The probabilistic conflict detection algorithm was evaluated during a pre-CE5 "shakedown" period. Conflicts were detected and pilots alerted through the aforementioned multi-leveled system. The escalation of alert levels from SA1 to SA3 followed a main evolutionary trend in the Probability-TLOS domain. This evolutionary trend is labeled as the "Main Sequence" in Figure 7. A very small number of alerts entered the main sequence midway from outside the shaded region. Those alerts manifested themselves as "pop-ups". Pop-ups were problematic in that they were likely already in alert level SA3 when they first appeared. This left the flight crew very little time to respond strategically.

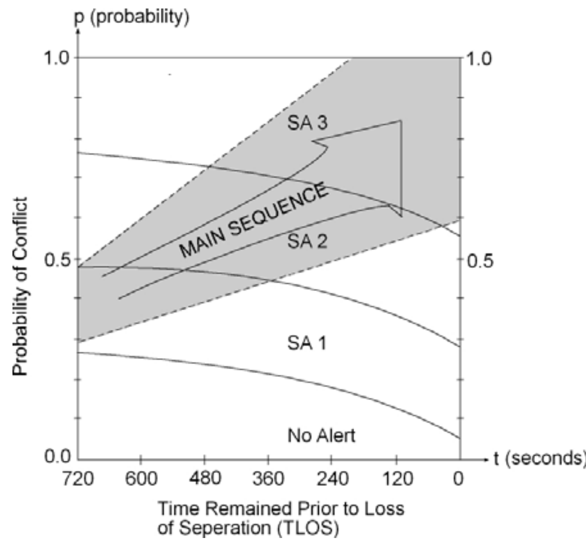


Figure 7. *Evolution of alert levels along the Main Sequence.*

Another artifact observed was alert dithering (i.e. a fluctuating alert level). This proved distracting to the pilots. Both pop-ups and dithering artifacts can be attributed to incompatible simulator behavior from different simulation platforms, the absence of a network wide time-synchronization system in the distributed simulation, and to a lesser extent the low density sampling of Monte Carlo space (500 iterations per cycle). Further study is needed in these areas.

A third artifact of the probabilistic algorithm was the violation of the aforementioned invariance. There were a very small number of cases in which the conflicting aircraft pair did not receive the same alert at precisely the same moment. This inconsistency was due to two probabilistic systems taking random samples independently (therefore, non-identical variance), as well as system messaging delays and the absence of a time-synchronization system. Further study is warranted in this area.

As an immediate remedy to these artifacts (and to further improve system performance), the conflict detection algorithm was re-configured to probe deterministically (by sampling the Monte Carlo space once per cycle). This was the version of the CD&R that went into the actual CE5 experiment.

Automated Conflict Resolution

The automated conflict resolution was implemented incrementally leading up to the pre-CE5 shakedown. While it worked well in simpler forms, its performance was less than ideal when more complex maneuver types were added to the solutions. The increase in complexity was compounded by the generation of new flight plans that were incompatible with other CDTI components. The result was less than ideal solutions and poor system performance. A decision was made to disable the automated conflict resolution feature for the actual CE5 experiment, and continue to resolve conflicts manually.

Rules of the Road Automation

ROR performed flawlessly during the shakedown and the actual CE5 experiment. It accurately applied rules and issued burdening settlements that could be consistently verified by the conflict aircraft pair. As a result, resolution maneuvers were made only by the burdened aircraft during autonomous-autonomous encounters, eliminating right-of-way ambiguities. Together with the deterministic conflict detection and alert, ROR fulfilled the role of the airborne self-separation tool for the autonomous flights during the CE5 experiment.

Conclusion and Future Work

Although the automated conflict resolution tool was not yet matured at the time of the CE5 experiment and had to be disabled, the overall CD&R-capable CDTI proved very successful. The Current Probe, the RAT Probe, the Vector Probe, and the ROR all contributed to enhancing the pilot's ability to resolve conflicts manually, a result consistent with previous work. It has been shown that pilot-generated resolutions are more effective when aided by decision support tools (Johnson, Bilimoria, Thomas, Lee, and Battiste, 2003).

While the concept and the design of the automated conflict resolution is sound, more work will be done to handle the complexity of multiple maneuver types and seamless interface with other CDTI components.

The dithering and the pop-up alert artifacts of the probabilistic conflict detection algorithm could be addressed with enhancements to the algorithm, the overall messaging system, and possibly with a denser Monte Carlo sampling. A new alert level mapping scheme should also be explored.

Finally, although the ROR performed flawlessly during the CE5 and handled all right-of-way issues, there was no provision in place to handle the case in which no rule applied. This case currently always defaults to burdening the ownship. So far, it has not occurred in experiments, but if it does, it will lead to the race condition because both aircraft will be burdened. A new ruling scheme is being developed for this special case.

References

Bilimoria, K.D. (2000). A Geometric Optimization Approach to Aircraft Conflict Resolution. *Paper No. 2000-4265, AIAA Guidance, Navigation, and Control Conference*, August 2000.

Johnson N.H., Canton R., Battiste V., Johnson W.W. (2005). Rules of the Road: Requirements and Implementation for a Simulation of En-route Self-Separation. Submission to the *2005 International Symposium on Aviation Psychology*, Oklahoma City, OK.

Johnson, W.W., Bilimoria, K.D., Thomas, L.C., Lee, H.Q., Battiste, V. (2003). Comparison of Pilot and Automation Generated Conflict Resolutions, *Paper No. 2003-5400, AIAA Guidance, Navigation, and Control Conference and Exhibit*, August 2003.

Prevot, T. (2002). Exploring the Many Perspectives of Distributed Air Traffic Management: The Multi Aircraft Control System MACS", *International Conference on Human-Computer Interaction in Aeronautics*, HCI-Aero 2002, Cambridge, MA, October, 2002.

Yang, L.C., Kuchar, J.K. (1997). Prototype Conflict Alerting Logic for Free Flight. *AIAA Journal of Guidance, Control, and Dynamics*, Vol. 20, No. 4, July-August, 1997.

Yang, L.C., Kuchar, J.K. (1998). Using Intent Information In Probabilistic Conflict Analysis. *AIAA-98-4237, AIAA Guidance, Navigation, and Control Conference*, Boston, MA, August 10-12, 1998.

INVESTING IN SAFER COMMUNICATIONS: PHRASEOLOGY AND INTELLIGIBILITY IN AIR TRAFFIC CONTROL

Ana Margarida Casaca
Instituto Superior de Psicologia Aplicada
Lisbon, Portugal

Teresa C. D'Oliveira
Instituto Superior de Psicologia Aplicada
Lisbon, Portugal
António Guerra
NAV, EPE
Lisbon, Portugal

Communications are central to air traffic control and any potential intervention that might contribute to its increased efficacy is considered relevant. This paper explores two main characteristics associated with communications: aeronautical phraseology and intelligibility. Although phraseology may contribute to an increased precision of the message, several factors may hinder it through speech intelligibility. In this study, air traffic controllers were asked to reproduce several messages that vary in phraseology correctness and speech intelligibility. Results suggest that considerable attention should be given to factors affecting speech intelligibility as increased numbers of errors and omissions were reported in messages with this characteristic.

Introduction

Communications critical role in Air Traffic Control is emphasized by its intervention in a variety of accidents and incidents (Davidson, Fischer & Orasanu, 2003). Although communications between controllers and pilots are standardized, errors may occur, and quite often with fatal consequences.
Communications between aircrews and air traffic controllers (ATCs) may be defined as the complete and effective transfer of information between these actors. This process of information transfer embraces many tasks and procedures that should be timely applied. In this process there are endless opportunities for human error to occur (Mackintosh, Lozito, McGann e Logsdon, 1999).

Hopkin (1995) argues that few studies on Air Traffic Control have actually considered the human factors associated with communications. Despite the lack of empirical support, phraseology and intelligibility are variables traditionally associated with communications' efficacy. The typical study considers the ATC as the sender of the information and the pilot as the receiver of such data. In this study, a different approach was used and ATCs were invited to analyze their colleagues' work. In particular, ATCs were asked to reproduce messages with varying degrees of correctness of aeronautical phraseology and speech intelligibility. This approach provided a unique opportunity to assess the controllers' awareness of the importance of communications to air safety in general and the potential implications of using non standard phraseology and unintelligible speech to the efficacy of aviation communications. As stated by Hopkin (1995), conveying information correctly actually represents a process that includes the listener's correct hearing and understanding. In Air Traffic Control, much of the richness of English and the flexibility and utility of speech must be curbed in the interests of standardization, intelligibility, completeness and the prevention of misunderstanding and error.

In this work the influence of professional experience on the ATCs ability to correctly reproduce a message was also analyzed.

Method

Participants

A total of 65 air traffic controllers, male and female, operating in the FIR (Flight Information Region) of Lisbon participated in this study. Volunteers were divided in three groups with distinct levels of professional experience: 12 ab-initio trainees, 11 less experienced controllers (up to 10 years of experience) and 42 very experienced controllers (more than 10 years of experience).

Instrument

A total of 30 real Air Traffic Control communications were recorded varying in phraseology (correct versus incorrect) and in intelligibility (intelligible versus non intelligible). Messages also varied in extension, the simpler ones with only two elements and the more complex ones with more than tem elements. Communications were assessed by an independent expert bearing in mind each of the aforementioned characteristics. In a within design, ATCs were asked to listen to the messages and to reproduce them in writing. Only one chance for listening was provided, a feature particularly relevant in air traffic control communications as the readback procedure is associated with greater efficacy.

Results

Results were analyzed considering the following issues: full reproduction of the message content; sequence of elements in transmission; partial reproduction of the message contents (i.e., only the general idea is recorded), omissions and their content; mistakes and their content.

Kruskal Wallis analyses suggested that significant differences were obtained for the four types of messages in what concerns textual reproduction [$\chi 2= 14.90$, df= 3, $p< .002$], sequence of elements [$\chi 2= 10.47$, df= 3, $p< .015$], general content [$\chi 2= 13.41$, df= 3, $p< .004$] and omissions [$\chi 2= 11.71$, df= 3, $p< .008$]. Best results were obtained for intelligible messages (regardless of phraseology) followed but messages with the correct use of phraseology and finally messages with incorrect use of phraseology and non intelligible. A two-way ANOVA revealed that only intelligibility is a significant factor for the number of errors [$F(3,1)= 5.531$, $p< .027$], with non intelligible messages being reproduced with more errors. A first implication of these results is the suggestion that intelligibility is the most important variable for communications' efficacy. The use of correct phraseology on its own is not a guarantee of greater efficacy in terms of communications.

Professional Experience

A clear distinction between ab-initio ATCs and ATCs with professional experience may be made with ab-initio ATCs presenting worst results in terms of textual reproduction, sequence of elements and general content. Best results were obtained for ATCs with up to 10 years of experience

Discussion

Results suggest that intelligibility plays a more central role in the reproduction of a message. If a message is intelligible, ATCs tend to reproduce it textually without errors or omissions regardless of the correct use of phraseology in the message. It is also important to emphasize that the lack of intelligibility in a message significantly increases the number of errors and omissions. One source of lower intelligibility is the presence of a specific accent of the sender speech which may cause ambiguities and doubts on the receptor regarding the message content. Such speech characteristics may disturb the efficacy of a given communication and therefore represent an avenue to improve the safety of the air traffic system. Familiarity with message content also represents an important issue. Difficulties in reproduction were evident in messages that included non-standard procedures or unusual aircraft call-signs even if the use of phraseology is correct and speech intelligibility is perfect.

In what concerns professional experience, the most important implication of this study regards the fact that ATCs with up to 10 years of professional experience present the best results in terms of message reproduction. Results regarding the effects of professional experience can be described as an inverted "U"relation. Ab-initio ATCs find it difficult to reproduce messages correctly. With experience, performance improves gradually. When a certain age limit is reached some breakdowns in the performance start to show, thus revealing professional experience does not always contribute in a positive way.

References

Davidson, J., Fisher, U. & Orasanu, J. (2003). *When language becomes a barrier instead of a bridge: Communications failures between pilots and air traffic controllers.* Paper presented at the 12th International Symposium on Aviation Psychology (April 14 - 17 2003). Dayton, Ohio.

Hopkin, V. D. (1995). *Human Factors in Air Traffic Control.* London: Taylor and Francis.

Mackintosh, M., Lozito, S., McGann, A., & Logsdon, E. (1999, October). *Designing Procedures for Controller – Pilot Data Link Communication: Effects of Textual Data Link on Information Transfer.* Paper presented at the World Aviation Conference, San Francisco.

A COMPARISON OF EVALUATIVE TECHNIQUES TO IMPROVE THE RELIABILITY OF MAINTENANCE DOCUMENTATION

Bonnie Lida Rogers[1], **Christopher J. Hamblin**[1], and **Alex Chaparro**[1,2]
National Institute for Aviation Research[1]
Department of Psychology[2]
Wichita State University
Wichita, Kansas

The purpose of this research was to investigate the applicability of usability methods in evaluating aviation maintenance documentation and to document the types of errors found. A diverse set of participants were recruited to participate in the evaluations in order to document how experience and training affect error detection. The results are similar to the findings of usability testing of software and web design – system experts and users identify unique errors and roadblocks.

Introduction

Maintenance procedures and information have been cited as primary factors contributing to maintenance errors (Dekker, 2002; Hobbs & Williamson, 2003; McDonald, Corrigan, Daly, & Cromie, 2000; Reason & Hobbs, 2003). A review of Naval Aviation Maintenance mishaps that occurred between 1990 and 2003 (Ricci, 2003) showed that 28% of the accidents involved problems in maintenance procedures including missing procedural steps, incorrect sequence of steps, inadequate procedures for inspection and troubleshooting, and incorrect technical information and diagrams. However, because mishaps are rare events, they underestimate the frequency of incidents in which poor documentation resulted in maintenance errors. Also, mishaps do not account for the other effects of poor documentation including the costs of incorrectly executed or slowed maintenance.

Maintenance documentation has recently begun to receive attention from academic researchers, the Federal Aviation Administration, and manufacturers. Many of these studies have focused on employing human factors principles to document and workcard design (Drury, Sarac, & Driscoll, 1997; Patankar & Kanki, 2001; Patel, Prabhu, & Drury, 1993). More recently, the methods and techniques employed by the aviation industry to develop maintenance documentation have also been investigated. Chaparro and Groff (2001) identified a number of problems with the development of maintenance documentation, including: reactive rather than proactive evaluation of the manuals, the limited use of aircraft maintenance technicians' (AMTs') input and procedure validation, the absence of systematic attempts to track error, and the lack of standards for measuring document quality.

In addition to improving maintenance documentation through design guidelines and manual usability, the accurate and clear communication of information is also critical. In other words, the AMT's interpretation of the procedure must match the intent of the writer for successful maintenance task completion. A mismatch has two likely outcomes. First, the AMT may become frustrated and call customer support for assistance in performing a procedure; or secondly, the AMT may "work-around" the procedure. The "work-around" approach entails trying to deduce the writers' intent when a procedure is confusing, or the information is incomplete or inaccurate.

This is not an uncommon occurrence. A study by Hobbs and Williamson (2000) conducted for the Australian Transportation Safety Bureau found that 67% of AMTs report having been misled by maintenance documentation, 47% report having opted to perform a maintenance procedure in a way they felt was superior to that described by the manual, and 73% of mechanics surveyed reported failing to refer to maintenance documents either occasionally or often. Chaparro, et al. (2002) also found that 64% of AMTs reported finding their own way of performing a procedure. Nearly 60% of AMTs reported continuation of an unfamiliar task despite not being sure if they were performing it correctly (Hobbs & Williamson, 2000). Similarly, McDonald et al. (2000) reported that 34% of routine maintenance tasks are performed in ways different than outlined in the maintenance documentation (MD).

Surveys reveal that aviation manufacturers rely on aircraft maintenance technicians (AMTs) to identify problems in MD (Chaparro et al., 2002). Most corrections to the MD are post-release through reports of problems by AMTs, called Publication Change Requests (PCRs). However, assuming that AMTs will report errors in maintenance procedures may be incorrect. Chaparro et al. (2002) found that 53% of AMTs reported only occasionally, rarely, or never reporting errors they found.

AMTs are often very good at deriving a plausible interpretation of incomplete information by drawing on their knowledge and that of other mechanics. This ability may result in an AMT misinterpreting procedures in such a manner that it is difficult to discover the error in their interpretation and subsequent actions. Although the AMTs' training and experience may allow them to correctly identify the writers intent, this will not always be the case. This uncertainty can be reduced by the proactive approach of assessing documentation quality *before* publication using tools originally developed to test the usability of computing software programs and documentation.

The purpose of these experiments is to investigate the applicability of two usability methods in evaluating aviation documentation and to document the types of errors found in MD. A diverse set of participants were recruited to participate in the evaluations in order to document how familiarity and training effect error detection.

Based on interviews with aviation technical writers, two usability techniques (described below) were chosen for the evaluation: Cognitive Walkthrough (CW) and User Performance (UP). Two experiments were performed to evaluate each of these evaluative methods.

Cognitive Walkthrough (CW) is a review technique in which evaluators review or "walk through" each step of a procedure to identify incorrect technical and factual information, poor wording choices, and inadequate information. Participants are instructed to visualize performance of each step as if they were doing the task. Normally, CW reviews are conducted in the early stages of document development to make corrections and changes before actual user testing.

User Performance Evaluation (UP) involves a participant physically performing a task. Participants are chosen who are not familiar with the task procedure or its development, to ensure that they are representative of users (AMTs) and the procedure can be evaluated without the potential biases arising from knowledge of the developer's, i.e. technical writer's, intent or familiarity with the system's design. Two forms of the UP were also compared: 1) a single user (SU) (i.e., AMT) performs the evaluation and 2) a two-person team work together, referred to as a Co-discovery (CD) user performance technique. In this study, an AMT performs the task as written in the MD and a Customer support engineer observes and makes comments.

Methods

CW Participants

Typically, CW evaluators are "expert" reviewers, familiar with the product's design and development; however, in this evaluation, we selected both "expert/familiar" and "naïve/ unfamiliar" participants to review the MD in order to investigate the role experience (expert vs. naïve) and training (AMT vs. engineer) play in error detection at earlier stages of document development.

Nineteen participants, 17 male and 2 female, completed the CW evaluation. The participants were assigned to one of four groups (expert vs. naïve) and technical background (engineers vs. AMTs). A total of three expert engineers, 5 expert AMTs, 6 naïve engineers, and 5 naïve AMTs participated in the evaluation. Naïve mechanics and engineers watched a short animated video of the procedure that illustrated the key parts and provided an overview of the task's process. One naïve engineer participant's responses were not included in the analysis as she reported more than the combined total of the other members in her group.

UP Participants

A total of ten naïve AMTs and five naïve engineers (all unfamiliar with the new procedural task) from the manufacturer's service facility participated in the UP Evaluations. Five of the AMTs were assigned to the single-user (SU) evaluation and five were assigned to the Co-discovery (CD) evaluation. The five naïve Customer Service engineers were teamed with the five naïve AMTs in the CD evaluations. All of the participants in this evaluation were male.

Materials

A general aviation aircraft manufacturer provided an unpublished maintenance procedure for the usability testing. This procedure was chosen because 1) it was unfamiliar to the pool of AMTs and their prior experience did not transfer readily to the new design, and 2) a computer simulation and physical prototype were available for use in testing. Prior to the experiments, the maintenance procedure was evaluated by production line mechanics and design engineers familiar with the task to estimate the number and types of errors within the document. The procedure was not modified as it was judged to have a sufficient number and types of errors.

CW Procedure

All participants read a paper copy of the MD and were asked to note any errors they found including typos, missing or incorrect information and any instructions that were out of sequence or unclear. Any materials typically referenced while proofing the MD (e.g. engineering drawings) were available to the participants while they reviewed the written procedure. The time required to complete the cognitive walkthrough was recorded upon completion (M = 40 minutes, range 26-70 minutes).

UP Procedure

AMTs were instructed to perform the procedure as written in the MD and to verbally describe what they were doing at each step and why they were doing it. In the CD evaluation, CS engineers were to observe. In the S and CD user performance evaluations, both types of participants (CS engineers and AMTs) were asked to verbalize their actions and inform the researchers of any instruction (or part of an instruction) that was incorrect, missing, out of sequence, or unclear. The time required to complete the cognitive walkthrough was recorded upon completion (M = 142 minutes, range 105-210 minutes).

Prior to the experiment all participants were informed of the purpose of the experiment and were asked to read and sign a consent form and privacy statement. Two researchers conducted the evaluations and recorded and coded the comments made by all participants into the error taxonomy, see Results section. A Cohen's Kappa (κ) of .85 was calculated on a sample of 50 comments reflecting an excellent level of consistency between the coders (Fleiss, 1981). Following the experiments, each participant completed a short background and satisfaction questionnaire.

Results

Error Taxonomy. To facilitate analysis and interpretation, a coding scheme was developed to categorize the errors identified by the participants. Within the context of this study, errors are defined as those items identified by participants as potential problem areas in the documentation. Four error-type categories and twelve specific reason categories were identified in the evaluations: 1) Technical (tools, values, parts); 2) Language (clarity of wording/terminology, grammar, typos, incorrect information); 3) Graphics (dimensions, part diagram, caption/text); and 4) Procedural (step(s), ordering). The associated corrective actions (add, delete, or change information) suggested by the participants' comments were also coded for analysis.

Cognitive Walkthrough (CW)

The results in Table 1 show that experts (AMTs and engineers) identified more than twice the errors (154 vs. 72) than their naïve counterparts. This is true despite the fact that there were fewer expert participants (n = 8 vs. 10). Both naïve and expert evaluators reported language error types most frequently (naïve, 41; expert, 63), followed by procedural-type errors (naïve, 19; expert, 47).

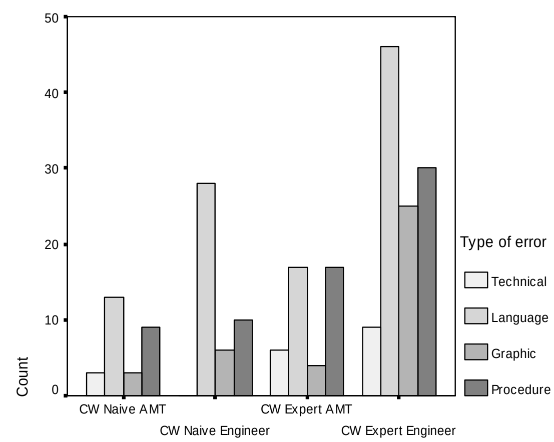


Table 1. *Number of errors reported in the CW method by evaluator group.*

A review of the comments made by each user group revealed several differences. Comments by naïve participants typically regarded the meaning or interpretation of the text and "what ifs?" (i.e., the absence of instructions regarding what actions to perform if a stated value or condition was not met.)

The experts reported more errors that were factual in nature including incorrect technical values, language, and procedural sequences. This result is not surprising since only individuals familiar (i.e., experts) with the design and operation can readily identify whether descriptive or factual information is incorrect.

These results illustrate the unique contributions made by the different experience (i.e., familiarity) levels of evaluators at an early stage of document development. Because of their familiarity with the procedure, system experts were better able to identify errors in technical information and system descriptions. However, due to their familiarity with the system they were less likely to identify vague, unclear, and imprecise procedural descriptions reported by the naïve participants.

User Performance Testing (UP)

AMT participants who were unfamiliar with the task and used the documentation to actually perform the procedure reported errors more frequently than any group in CW or the Customer Service Engineers in the Co-discovery (CD) method of user performance testing. The CD evaluation method was relatively more effective in identifying errors than the SU method – roughly twice as many total errors were reported by participants using the CD vs. the SU method (CD, 331; SU, 162).

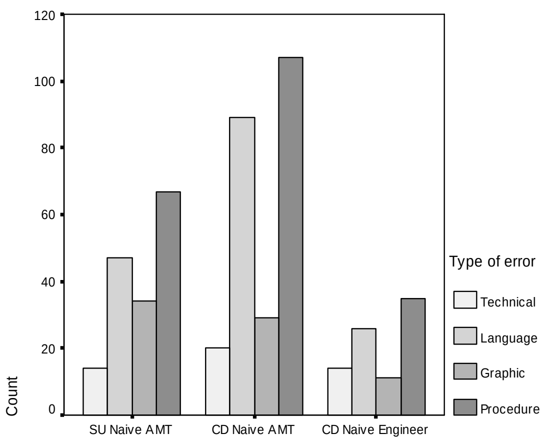


Figure 2. *Number of errors reported in SU and CD user performance methods by evaluator group.*

A comparison of the contributions made by AMTs and engineers in the CD method show that AMTs identified many more errors (roughly three times) associated with procedural and language than did the engineers. Like the results from the CW, procedural and language errors were again the most frequently cited problems. The most common types of procedural errors were missing steps including the absence of instructions regarding what actions to perform if a stated value or condition was not met, steps for disassembling or reassembling, and simple instructions which aid the AMT frame of reference (e.g., open/close door).

Comparison of CW and UP. Figure 3 illustrates the average number of the four major error types (language, graphic, procedural and technical) reported by participants using the two evaluation methods (CW and UP). These results demonstrate the benefits of performing the maintenance procedure on an aircraft. As illustrated in the differences between the frequency of language and procedural errors in Figure 3, the CW was relatively more effective at detecting language errors while the UP evaluations resulted in greater detection of procedural errors.

A comparison of the specific reasons the error was reported reveals that the UP evaluations were effective in spotting language errors related almost exclusively to clarity; whereas, the CW technique identified a more diverse set of language errors including grammar and typographic errors. Incorrect information was found most frequently by expert evaluators in the CW but was also reported by naïve participants in UP testing.

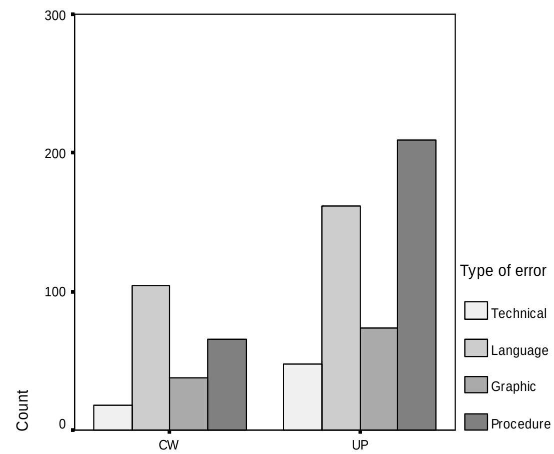


Figure 3. *Error frequencies as a function of evaluation method and error type.*

Procedural errors identified in UP evaluations were most frequently missing steps (n = 95), followed by the need to change the sequencing of the steps (n = 44). Both of these specific reasons were reported more than three times as often in UP as in CW.

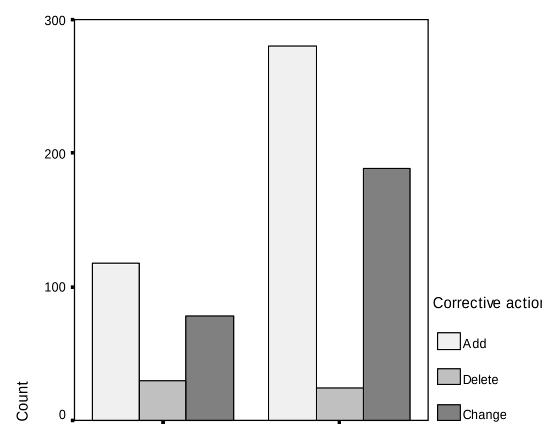


Figure 4. *Comparison of the corrective actions by evaluation method.*

Corrective actions of adding, deleting and changing information were implied when the errors were reported. As illustrated in Figure 4, the majority of these comments for both User Performance (SU and

CD) and Cognitive Walkthrough (CW) techniques requested either changing or adding more information to the procedure. Note that more than twice as many comments requesting that information be added to the procedures were obtained through UP (n = 280) than CW (n = 118).

Unique Errors. In many instances, the same error was reported by more than one participant in the experiment; these redundant reports were eliminated and the sums of these single instance or "unique" errors for each method were calculated. Sixty-seven percent of the 226 reported errors in CW and forty-four percent of 493 in UP were unique errors. This analysis also shows that the two techniques (i.e., CW and UP) were not redundant as the CW method had only 21 errors in common with the SU and 45 errors in common with the CD method.

Satisfaction Measures. A scale was developed to assess the participants' satisfaction with the written procedure and was administered following the usability testing. The scale had ten individual statements of satisfaction measured on a 5-point agreement scale; Strongly Disagree to Strongly Agree. A Cronbach's Alpha of .92 was calculated revealing excellent scale reliability in measuring participants' satisfaction with the technical documentation (Nunnally, 1978). Three additional statements asked for 1) a judgment of the procedure's complexity relative to other procedures; 2) whether additional instructions would be needed to complete the procedure; and 3) an open-ended query of what would improve the procedure. Results of the satisfaction measures were analyzed by method, i.e., CW and UP evaluations (Single-User (SU) & Co-Discovery (CD), and by user group, (expert engineer, expert AMT, naïve engineer, and naïve AMT).

Generally, participants in the CW method were more satisfied with the written procedure, giving it a mean rating of 3 or higher (i.e., greater satisfaction) on the ten satisfaction statements and the overall satisfaction query; whereas, those who participated in UP evaluations rated the procedure <3, (less satisfaction) for those statements. The total satisfaction score for the CW group (M = 68.33, SD = 14.86) was significantly higher than for the UP group (M = 54.00, SD = 14.38), $t(31)$ = 2.79, p = .009, d = .93. Responses for the following satisfaction queries were significantly higher for the CW participants in comparison to those tested by UP: "I am satisfied with the number of steps included." (CW: M = 3.78, SD = 1.11; UP: M = 2.33, SD = .90), $t(31)$ = 4.04, p = .001; "The procedure was clearly written." (CW: M = 3.47, SD = .96; UP: M = 2.40, SD = .83), $t(31)$ = 3.43, p = .002; "The illustration was helpful." (CW: M = 3.68 , SD = 1.64; UP: M = 2.80 , SD = 1.27), $t(31)$ = 2.08, p = .046; and "The amount of information included was useful." (CW: M = 3.78, SD = 1.00; UP: M = 3.20, SD = .90), $t(31)$ = 1.98, p = .056. Both groups indicated that the procedure needed more instructions and were neutral that this procedure was "more complex than most."

A comparison of the number of errors reported and satisfaction score reveals that satisfaction scores are negatively related to the number of errors found – as the number of errors reported increases, the level of satisfaction significantly decreased (r = -.66, p < .01).

Discussion

The results of this investigation show that 1) User Performance and Cognitive Walkthrough evaluations are complementary techniques for evaluating maintenance documentation, 2) the errors identified by individual participants varied in significant ways according to familiarity (expert vs. naive) and training (engineers vs. AMTs), 3) procedural and language errors are the most commonly cited errors reported in the maintenance documentation usability testing., and 4) satisfaction levels are higher in a CW evaluation compared to UP evaluations.

Cognitive Walkthrough (CW). Most commonly in usability evaluations, the user does not review the task at the early stage of development; however, results from this study show that in this domain (i.e., aviation maintenance), the information from a naïve user (AMT) and naïve engineer may provide the technical writer with valuable feedback as to what areas may need additional clarity and where procedural steps, such as checks and functional tests may need to be added.

Several issues identified by the naïve participants in the CW were later reported as problems in the UP evaluation. For example, three naïve engineers in CW testing reported that the wording "Adjust …until the force needed to close … are the best between them." needed clarification. In the UP evaluations, this step was also cited as unclear by three of the naïve AMTs in SU evaluations, two naïve engineers and two naïve AMTs in the CD evaluations. When it is not possible to test MD using a UP, CW may be a viable alternative using naïve users (AMTs) and naïve engineers for evaluations.

User Performance (UP). User performance testing identified specific areas in the documentation that were incomplete, unclear, or incorrect. Ambiguities are more salient to the user when they have to convert written statements into action. In addition, physical obstructions that make the procedure difficult or impossible to perform become obvious. The results also demonstrate the benefits derived from having evaluators work as a team.

Problems with language clarity included the use of unfamiliar part names, lack of consistency in the procedure, and subjective language, such as "…seal can be removed." As one AMT commented, "Does it need to be removed or not?" Another statement in the procedure was "make sure … operates correctly" to which an AMT commented, "What is correctly? Correct gap or correct position?). When unfamiliar part names were referenced, the AMTs would often rely on their experience to identify the relevant part. This was not always sufficient as several of the AMTs volunteered that they would have taken apart or adjusted the wrong component.

Given that the same types of information obtained in usability evaluations of MD are similar to those cited as contributory to accidents and incidents (Ricci, 2003), it would seem that adapting usability techniques to improve MD is a feasible and proactive alternative to the current MD development methods. The two methods tested in this research yielded a significant number of instances in which both inaccurate and unclear information could be corrected before publication.

Additional benefits to employing these methods include increasing technical writers' awareness of the information necessary for the AMT to perform maintenance and a consideration of the constraints under which the AMT is working. As part of this research, an aviation technical writer's "toolbox" was developed that outlines evaluative methods which have been adapted for aviation technical documentation. The toolbox consists of descriptions of each evaluation technique, guidelines for using the methods, and various supporting documents (questionnaires, data collection forms, etc.) that can be used during the evaluations. The toolbox is available at http://www.niar.wichita.edu/humanfactors/toolbox/default.htm .

References

Chaparro, A., & Groff, L. S. (2001). *Human factors survey of aviation technical manuals phase I report: Manual development procedures* (No. DOT/FAA/AR-01/43). Washington, D.C.: Federal Aviation Administration.

Chaparro, A., Groff, L. S., Chaparro, B. S., & Scarlett, D. (2002). *Phase 2 Report: User evaluation of maintenance documents* (No. DOT/FAA/AR-02/34). Washington, D.C.: Federal Aviation Administration.

Dekker, S. (2002). *The Field Guide to Human Error Investigations*. Aldershot, U.K.: Ashgate Publishing Ltd.

Drury, C. G., Sarac, A., & Driscoll, D. M. (1997). *Documentation design aid development.* Washington, D.C.: FAA Office of Aviation Medicine.

Hobbs, A., & Williamson, A. (2000). *Aircraft maintenance safety survey - Results*. Canberra: Australian Transport Safety Bureau.

Hobbs, A., & Williamson, A. (2003). Associations between errors and contributing factors in aircraft maintenance. *Human Factors, 45*(2), 186-201.

McDonald, N., Corrigan, S., Daly, C., & Cromie, S. (2000). Safety management systems and safety culture in aircraft maintenance organisations. *Safety Science, 34*, 151-176.

Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill.

Patankar, K., & Kanki, B. G. (2001). *Document Design Strategies for Improving Airline Maintenance Procedures.* Paper presented at the 11th International Symposium on Aviation Psychology, Columbus, Ohio.

Patel, S., Prabhu, P. V., & Drury, C. G. (1993). *Design of workcards.* (No. DOT/FAA/AM-93/15). Springfield, VA: National Technical Information Service.

Reason, J., & Hobbs, A. (2003). *Managing Maintenance Error*. Aldershot, U.K.: Ashgate Publishing Ltd.

Ricci, K. (2003). *Human factors issues in maintenance publications design.* Paper presented at the DOD Maintenance Symposium & Exhibition, King of Prussia, PA.

AUTOMATION, CRM AND DISTRIBUTED COGNITION: AN EXPLORATION OF THE DEFENSE MECHANISM IN THE COCKPIT

Jingnong Chen, **Douglas Paluszak** and **Julie Silverstein**
State University of New York at Buffalo
Buffalo, NY

What is the mechanism that allows aircraft flight crews to achieve such an astounding safety record despite the hazards they encounter? In this paper, we discussed the topics of aviation safety from a broad theoretical framework, which generally relate to these three topics: Automation, Crew Resource Management, and Distributed Cognition. We outline the preliminary results of a study surveying 38 reports from the Aviation Safety Reporting System. In this survey, the reports were given three classifications, the problem-based classification, the optimal-solution-based classification and the actual-solution-based classification. Some interesting findings were shown by studying the correspondences between three classifications. Based on the findings, an integrated defense mechanism with the contributions of automation, CRM, and distributed cognition was explained against the external and internal threats found in an aircraft cockpit.

Introduction

In 2000, 629 million passengers boarded airplanes at U.S. airports, yet the number of fatalities reported from passenger aircrafts accidents was approximately two hundred. Generally, aviation is considered a highly complex activity, with a hazardous and multifaceted threat environment; yet, air carriers consistently operate at a high level of reliability and safety. Why? What is the mechanism that allows aircraft flight crews to achieve such an astounding safety record despite the hazards? In previous studies, experts discussed the topics of aviation safety from a broad theoretical framework, which generally relate to these three topics: Automation, Crew Resource Management, and Distributed Cognition.

Automation

Automation involves the substitution of automation components for tasks that the machine may perform more efficiently than humans, or tasks which humans are incapable of performing safely or at all (Wiener, 1985). Cockpit automation is a typical example of a complex control environment. Every day thousands of flight crewmembers operate aircraft utilizing a variety of automated devices. These devices include everything from traditional autopilots and flight directors to elaborate flight management systems, aircraft performance management systems, and a host of automatic warning and alerting systems.

In the quest for safer and more efficient flight, microprocessor technology has enabled the rapid advance of cockpit automation, the principal rationale being the assumption that the reduction of the flight crew’s routine tasks and mental cognitive activities will reduce potential problems in the cockpit (Sarter & woods, 1994). This allows more time to supervise the flight operations effectively. Cockpit designers are incorporating more and more automation into the cockpit in an attempt to address human limitations; with their ultimate goal of automating the hazards out of the cockpit.

Overall, the movement toward cockpit automation has undoubtedly enhanced aviation safety, however to some extent it has become evident that automation doesn’t always replace the pilot’s in the cockpit, instead it changes the nature of their tasks, and therefore new sources of cockpit error have been created (Parasuraman & Riley, 1997).

Crew Resource Management (CRM)

Personnel-related causes or factors were cited in 89.8% of all general aviation accident reports for 1999 (NTSB, 2003). This realization led to the development of many programs that are used to improve what is called crew resource management (CRM). These programs aim at preventing aviation accidents by enhancing team performance through training. However so far, CRM is not defined explicitly. More generally Salas, Prince, Bowers, et al. (1999) conceptualized CRM as a “family of instructional strategies that seek to improve teamwork in the cockpit by applying well-tested training tools (e.g., simulators, lectures, videos) targeted at specific content (i.e., teamwork knowledge, skills and attitudes)”.

Because of this diversity, there are widely varying ideas about what constitutes CRM throughout the aviation community. Some CRM focused heavily on attitudes toward teamwork, pilot personality, and social interactions. Other programs focused mainly on behavior skills. As such, different labels, descriptions, and representations are used to define those skills. Evidence of the effectiveness of CRM training was obtained by many researches. For example, Continental Airlines’ Error Management training program, which is a CRM training program, was an effective accident prevention tool for helping cockpit crew identify, respond to, and resolve mistakes before they become a threat to flight safety.

Although much progress was made in the previous CRM training applications and researches, there are still some topics needed to be explored. This will require further study of the cognitive processes underlying team situation assessment, team situation awareness, and team decision-making, and the theoretically driven and practically relevant principles, guidelines, interventions, and tools are still much-needed resources.

Distributed cognition

Distributed cognition is an important socio-psychological phenomenon of the safety system in the cockpit (Hutchins & Klausen, 1990). Three properties of distributed cognition make valuable contributions to aviation safety. First, the overlapping communication makes the storage and dissemination of information flexible, in that the efficiency of receiving and transferring information is not only influenced by the personal skill and expertise, but it also utilizes the crew’s capacity to share information in the distributed networks.

Second, the creation of artifacts driven by distributed cognition is another practical contribution to aviation safety. In the advent of new technology, a significant number of powerful external symbolic devices and material memoranda are designed. Distributed cognition is viewed as the interactions between internal and external representational structures. In the cockpit, increasingly more information is arranged by an external representational structure, which is designed to conserve the limited resources in human working memory.

Distributed cognition’s third contribution to aviation safety is that its propagation reconstructs the cockpit culture on a deeper cognitive level that can be seen as an overall improvement in the level of situational awareness and aviation safety. It takes a culturally constituted functional group as its unit of analysis, rather than an individual mind. In doing so, aviation safety and efficiency are fostered in terms of breaking through the individual constraints and generating the positive behavior pattern of cooperation and coordination among the flight crew. As such, the flight crew as a whole has a greater awareness than the sum of its parts.

Initial thought of a defense mechanism in the cockpit

So far the contributions of automation, CRM, and distributed cognition to the aviation safety have been discussed separately. There has been no model or theory proposed to integrate all these contributions simultaneously. In this study, we are interested in exploring a defense mechanism against threats to safety in the cockpit, by integrating all three coping strategies driven by automation, CRM, and distributed cognition. Due to the diversity of their contributions in detecting and solving problems in the cockpit, we speculate that the efficiency of this kind of defense mechanism will be greater than those concerned with only single coping strategy. The benefit of this defense mechanism is to show how each coping strategy mutually supports the others by overcoming the limitations of each, which are discussed in this section.

Survey

Database and analysis

The method of this study was an archival data analysis. The data used in this study was obtained from the 50 ASRS reports found in the CRM Database Report Set dated October 9, 2003. To decompose these reports into meaningful classifications, five reviewers reviewed each case separately, and then discussed all fifty cases as a group. By consensus, the reviewers decided that 38 of these cases provided sufficient information for subsequent analysis. Each of these 38 cases used were analyzed according to three classifications which are discussed below. By studying the correspondences between three classifications, the defense mechanism with the contributions from automation, CRM, and distributed cognition was explained. Descriptive statistics were used to get some interesting findings

Three classifications

In each of the 38 cases, the incidents as reported by the flight crewmembers were given three classifications. The first classification - the problem-based classification – which defined the two groups of operational areas that caused the majority of the incidents reported in the ASRS Database. These two groups are the human performance errors (HPE) or external physical threats (EPT) and were proposed in studies by Shappell and Wiegmann (2004), and Gordon, Flin, and Mearns (2001). Within each group the types of problem were defined as the following (see Table 1).

Table 1. *The first classification*

General groups	types of Problem
HUMAN PERFORMANCE ERRORS (HPE)	Tactical decision error
	Perceptual error
	Communication failure
	Violations
	Misuse of Procedures
	Manual control failure
	Misuse of Checklists
EXTERNAL PHYSICAL THREATS (EPT)	Environment
	Weather
	Airspace structure
	Aircraft
	Maintenance
	Others

The second classification - the optimal-solution-based classification - is based on how the problems in the ASRS reports surveyed should have solved as reported by flight crewmembers. As proposed in the previous section, Automation, CRM, and distributed cognition each had different characteristics that benefit aviation safety, and their classification criteria were defined based on the following characteristics:

(1) Flight crew should have used automation to decrease workload and stress therein to solve the specific problem in the selected case (optimal-AUTO);

(2) Flight crew should have used CRM skills and strategies enhanced by training, minimize the resource expenditure and eliminate the human error therein to solve the specific problem in the selected case (optimal-CRM);

(3) Flight crew should have used an efficient distribution of cognitive activities throughout the cockpit to increase the information redundancy therein to solve the specific problem in the selected case (optimal-DC);

(4) Flight crew should have used other strategies to solve the specific problem in the selected case (optimal-other).

The third classification - the actual-solution-based classification - is based on how the problems in the ASRS reports surveyed were actually solved as reported by the flight crewmembers. This classification maps directly to the second classification, and its classification criteria is defined as follows:

(1) The flight crewmembers actually used automation to decrease workload and stress therein to solve the specific problem in the selected case (actual-AUTO);

(2) The flight crewmembers actually used CRM skills and strategies enhanced by training to minimize the resource expenditure and eliminate the human error therein to solve the specific problem in the selected case (actual-CRM);

(3) The flight crewmembers actually used an efficient distribution of cognitive activities throughout the cockpit to increase the information redundancy therein to solve the specific problem in the selected case (actual-DC);

(4) The flight crewmembers actually used other strategies to solve the specific problem in the selected case, or the problem was not solved (actual-other).

Results

The correspondence between the problem-based and optimal-solution-based classifications

Figure 1 shows the correspondence between the first, the problem-based classification, and the second, the optimal-solution-based classification. It revealed that HPEs, as well as EPTs, were extensively distributed throughout all levels of optimization-based solution classification. Overall, a review of the ASRS reports in the survey showed there were slightly fewer problems involving EPTs than HPEs, except that the number of HPEs that should have been solved by using an efficient distribution of cognitive activities relatively was higher than the number of EPTs. Furthermore, there were more EPTs that should have been solved by strategies other than automation, CRM and distributed cognition, than HPEs that should have been solved by these same strategies. However, in total the unsolved problems were much fewer than those that were solved.

Figure 1: The correspondence between the problem-based and optimal-solution-based classifications

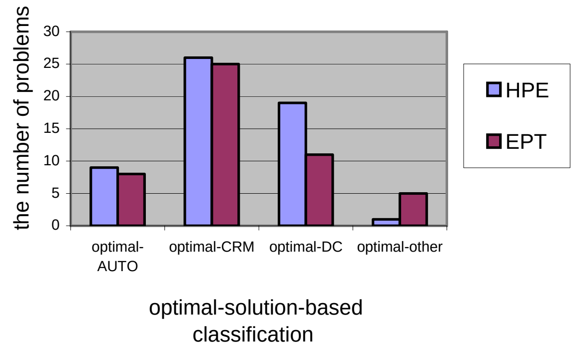
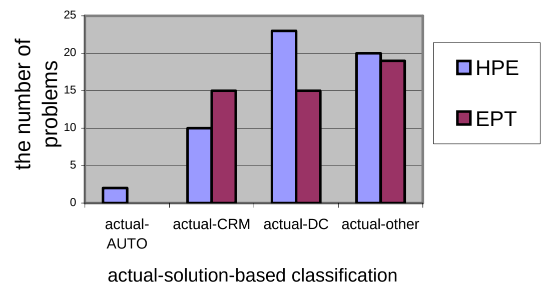


Figure 2: The correspondence between the problem-based and actual-solution-based classfications



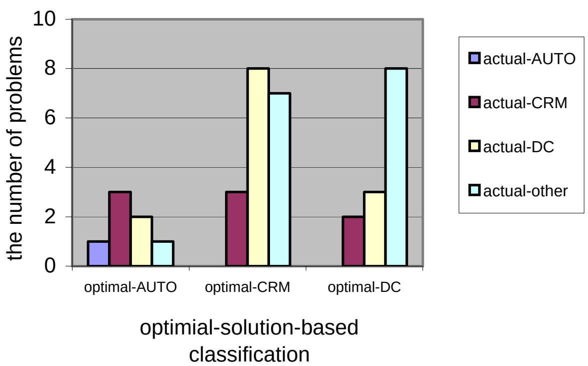
The correspondence between the problem-based and actual-solution-based classifications

Figure 2 shows the correspondence between the first, the problem-based classification, and the third, the actual-solution-based classification. Only two problems were solved by using automation. Meanwhile more HPEs were solved by using an efficient distribution of cognitive activities, than by using CRM strategies and skills. However, when the EPTs were eliminated, the benefits from distributed cognition were not different then those from CRM. These findings suggested that distributed cognition was powerful enough to solve more HPEs rather than EPTs, but CRM appeared to be effective in solving as many HPEs as EPTs. On the other hand, a certain amount of HPEs, as well as EPTs were solved by chance or even worst, were not solved.

The correspondence between the optimal-solution-based and actual-solution-based classifications

Figure 3 shows the correspondence between the second, the optimal-solution-based classification, and the third, the actual-solution-based classification. Most problems that should have been solved by using automation were solved by the use of effective CRM skills and strategies; most problems that should have been solved by effective CRM skills and strategies were solved by the flight crew operating at a high level of distributed cognition; and finally, most problems that should have been solved by distributed cognition were solved by using some other strategy, than automation, CRM, or distributed cognition; or they were not solved at all.

Figure 3: The correspondence between the actual-solution-based and optimal-solution-based classifications



It should be noted that if any problem was actually solved by the optimal strategies, these problems might not have been reported to ASRS. Therefore it is not always possible to determine how many incidents were actually solved by the flight crewmembers using the optimal strategy, from review of the ASRS reports alone. Except for the problems which were actually solved by the optimal solutions, figure 3 shows the distribution of the alternatives of the optimal coping strategy which solved the problems. Moreover, the proportion of these problems classified as "actual-other" formed a continuum, with the least number occurring when they should be solved by using automation, followed by CRM, and then distributed cognition.

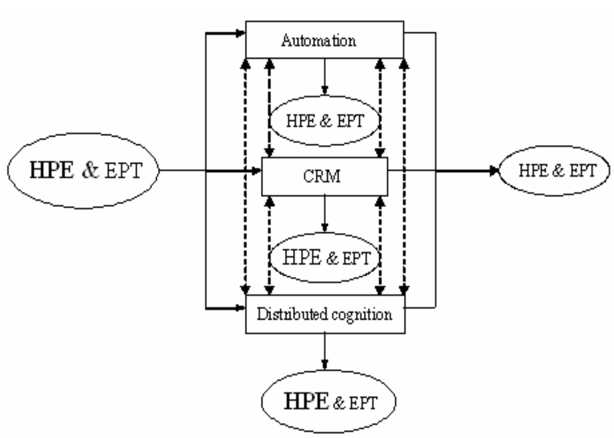
Discussion

Based on the correspondences among the three classifications, some interesting findings were shown to be relevant to our initial thought of an integrated defense mechanism in the cockpit. Overall, we believe that the power of a defense mechanism supported by automation, CRM and distributed cognition is strong. Almost 95% of problems in the cockpit could be solved by at least one of these coping strategies. Meanwhile, the three coping strategies are more effective to detect and solve HPEs than EPTs, especially for distributed cognition. Human errors caused nearly 80 percent of corporate aviation accidents during 1992-1997 (Hinson, 1997). Therefore, it is possible that this integrated defense mechanism could address the threats that cause the most types of aviation accidents.

The most important finding of this study is the correspondence between the second, the optimal-solution-based classification, and the third, the actual-solution-based classification. Except for the problems which were actually solved by the optimal solutions, most problems that should have been solved by using automation were solved by the use of effective CRM skills and strategies; most problems that should have been solved by effective CRM skills and strategies, were solved by the flight crew operating at a high level of distributed cognition; and finally, most problems that should have been solved by distributed cognition were solved by using some strategy, other than automation, CRM, or distributed cognition; or was not solved at all. This suggests that the coping strategies driven by automation, CRM, and distributed cognition, not only contribute to aviation safety individually, but may also compensate for the limitations of the other strategies.

Based on the findings above, we purpose a simple model of our integrated defense mechanism (see Figure 4). In this model, the issues represent the results of the survey in this study. The circles in the left and in the right represent the input of problems and the output of unsolved problems respectively. The links between the circle in the left and the squares demonstrate that the flight crews choose the appropriate coping strategies for the problems. In the middle, the links between the circles and squares represent the throughput of each coping strategy and the dashed links between the squares show the redundancy in the integrated defense mechanism; in that even if the flight crews does not choose the appropriate coping strategies for the problems, it still can be solved by using other strategies.

Figure 4: An integrated defensive mechanism in the cockpit



Notes 5: The circle and square represent the problem and the coping component respectively. The font of the words in the circle represents the amount of the problems.

In fact, the results showed that the flight crewmembers did not always recognize the benefits of the integrated defense mechanism as they usually did not use the appropriate coping strategies to solve the problems, or it went unsolved. This fact suggests that right now, our integrated defense mechanism is not being utilized efficiently to maximize aviation safety. There is a gulf between what the optimal solution was and what the flight crewmembers actually use to solve the problems. So, what can we do to eliminate this gulf, and thus increase aviation safety? Based on our findings, we recommend that to maximize the benefit of the defense mechanism in each cockpit, the coping strategies driven by automation, CRM and distributed cognition should be considered simultaneously to an increase in aviation safety, and further research into this theory is needed.

References:

Gordon, R., Flin, R., & Mearns, K. (2001). Designing a human factors investigation tool to improve the quality of safety reporting. *Proceedings of the human factors and ergonomics society 45th annual meeting*, 1519-1523.

Hinson, D. (1997). Teamwork and improved communication between people. *Scientific American*, 276(5), 62-66.

Hutchins, E., & Klausen, T. (1990). Distributed cognition in an airline cockpit. *Cockpit Cognition*, 1-32.

National Transportation Safety Board (2003). Annual review of aircraft accident data. *U.S. General Aviation, Calendar Year 1999.*

Parasuraman R., & Riley V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human factors*, 39(2), 230-253.

Salas, E., Prince, C., Bowers, C.A., Stout, R.J., Oser, R.L., & Cannon-Bowers, J.A. (1999). A methodology for enhancing crew resource management training. *Human Factors*, 41(1), 161-172.

Sarter, N. and Woods, D. D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system. *International Journal of Aviation Psychology*, 4, 1-28.

Shappell, S., and Wiegmann, D. (2004). Human error and general aviation accidents: A comprehensive, fine-grained analysis using the human factors analysis and classification system. *Building on the legacy, Fiscal Year 2003 Report*, Federal Aviation Administration's Human Factors Research & Engineering Division, 43-44.

Wiener, E. L. (1985). Beyond the sterile cockpit. *Human Factors*, 27(1), 75-90.

A COMPARISON OF SAINT WITH IMPRINT AND MICRO SAINT SHARP

Gerald P. Chubb
The Ohio State University
Columbus, Ohio

SAINT was a hybrid modeling and simulation language developed in FORTRAN for main frame computers that allowed simulation of human activities in the context of system operation. MicroSaint was initially developed in the C language, specifically for Personal Computers (PCs), mimicking much but not all of what was in the original FORTRAN version. IMPRINT Version 7 uses Micro Saint IV as its underlying computational engine. MicroSaint Sharp is based on the C# programming language and will be the computational engine underlying IMPRINT Version 8. Representational capabilities of these various modeling techniques are compared to illustrate what improvements have been made and what has been abandoned in the progressive development of these analysis tools.

Introduction

SAINT is an acronym standing for Systems Analysis of Integrated Networks of Tasks. As a modeling tool, it uses a general activity network to represent procedural and decision making tasks, including parallel activity chains by one or more operators. The associated software executes a Monte Carlo simulation of the activity network, generating statistics on activity duration, time of task sequence completion, number of task repetitions, and other descriptive measures. There are numerous similar diagramming techniques, such as: DeMarco Data Flow Diagrams (Yourdan and Constantine, 1979), Petri Nets (e.g., Desrochers and Al-Jaar, 1995), and PERT charts (Moder, et al., 1983). While PERT uses an activity-on-branch representation, SAINT uses an activity-on-node representation.

Background

The original impetus for developing SAINT was the Siegel-Wolf two-man operator simulation model (Siegel and Wolf, 1969), used in a study of F-106 nuclear vulnerability / survivability (Chubb, 1971). Task times were assumed to be normally distributed with some specified probability of success. Failed tasks led to repetition of the task. Average and standard deviations of the nominal task durations were adjusted to reflect the impact of time stress, as determined from time available versus time required. It was recognized that: 1) engineers were reluctant to use a model developed by psychologists, 2) there were other distributions of task times that might better represent certain activities, 3) the branching structure logic was simplistic, and 4) there was no representation of system dynamics that might drive human performance. To be effective, it was believed the best approach was to use simulation technology that engineers were taught to use and then incorporate human factors considerations into that technology. Such was the goal for SAINT.

SAINT Development

SAINT was developed using elements of GERT (Pritsker and Happ, 1966 and Pritsker and Whitehouse, 1966), a FORTRAN simulation language used by industrial engineers to model discrete systems, later adding elements from GASP IV (Pritsker and Hurst, 1973) that also allowed representation of continuous system dynamics (Cellier,1982). For a more detailed overview of the initial SAINT modeling and simulation capabilities, a list of preliminary applications, and references to documentation see Seifert and Chubb (1978).

Subsequent Applications and Developments included modeling of the B-1A Electronically Agile Radar (EAR) to determining the characteristics of time-sharing between forward looking terrain tracing and horizontal ground mapping modes. Additional branching logic and other modeling improvements were also made under the Cockpit Automation Technologies (CAT) program (Hoyland, et al., 1988). SADT (Marca and McGowan, 1988) was also shown to provide a good top-down, front-end analysis technique consistent with later developing the SAINT activity networks (Chubb, 1989).

Deficiencies and Shortcomings of SAINT included a lack of graphics capability to represent network models, complicated symbology for network diagramming, particularly the types of branching logic, and the general lack of technical support. While the source code was delivered with no restrictions on data rights (therefore available in the 'public domain' and releasable to any requester), there were no formal provisions for giving users any technical support if they encountered difficulties in their use of SAINT. SAINT was designed as a batch program for a large, main frame computer, and all data was in alphanumeric form using punched cards. Micro Saint (Laughery, 1985) changed that, substantially.

Micro Saint Development

Micro Saint was developed by Micro Analysis and Design (MA&D) as part of the Army's Chemical Defense program and made four major improvements: 1) it was hosted on personal computers (PCs), b) it included graphical representations of the network model, 3) it simplified the representation of branching logic, and 4) data entry was easier and less error prone. Micro Saint was initially programmed in the C programming language, which permitted capabilities not readily available to FORTRAN programmers. Network graphics were animated to indicate which task(s) were being executed as the sequence unfolded. In addition to helping users better understand the task-flow, it also helped 'debug' incorrectly implemented models. There was no attempt to include all the distribution types or types of branching logic found in SAINT, nor did Micro Saint include SAINT's continuous-time modeling of system dynamics.

The most recent version of Micro Saint is based on C#, not C or C++. This new language offers more programming power and additional capabilities (Bloechle and Schunk, 2003). The ability to build enhanced animation of models has also been added, as well as permitting the development of better web-based applications. However, developing advanced animations may, by itself, take as long as developing the model and Micro Saint simulation. It has a distinct cosmetic advantage in promoting a model and its use, but does little technically – the underlying model is the same. An add-on optimization package is also available for analyzing the output from a series of model runs.

Micro Saint is a proprietary product and therefore not in the public domain. However, MA&D does offer an academic discount for both student and the 'industrial strength' versions of Micro Saint. They also provide excellent training in their product (as well as appropriate courses in the use of IMPRINT).

IMPRINT Development

IMPRINT (Anonymous, 2003 a & b) is a tool developed by MA&D for the Army that helps satisfy part of the MANPRINT requirements Booher (1990). Version 7 uses Micro Saint IV as its underlying computational engine. Version 8, currently under development, will use Micro Saint Sharp as its underlying computational engine, providing some new / enhanced modeling capabilities for IMPRINT that are not treated in this comparison. The Army prefers Law and Kelton (2000) as their basic reference text on simulation and modeling.

IMPRINT has its own graphical user interface and may be used to look at both operator (e.g., individual missions) and maintainer (e.g., sustained combat operation) applications. There are now three levels or modes of modeling that are increasingly complicated and demanding of user input data. All three have 'standard' outputs built-in.

The simplest model implementation permits workload assessments of hierarchical task network models using the McCracken-Aldrich model (1984). The advanced workload assessment mode, restricts the modeling to a single level of task representation, but allows parallel tasking and uses the North model of workload (1989).

The most complex use of IMPRINT uses techniques originally developed under the CART program (e.g., Brett, et al., 2002). This permits goal-directed task modeling that better represents the way in which most missions are accomplished. More recently, the interface of IMPRINT and ACT-R has been explored as well (Kelley and Scribner, 2003).

Both the advanced workload and CART-related modes of IMPRINT give the user access to more powerful modeling tools but fall short of requiring a complete understanding of the full complexities of Micro Saint. This structured support of increasingly more complex models allows new users to systematically develop their modeling expertise.

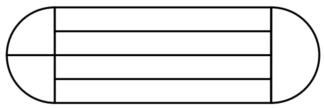
While the documentation does not completely support the user's needs, the Army has made provision to give technical support to new users, something the Air Force did not do for SAINT. This substantially enhances IMPRINT's utility. IMPRINT is a non-proprietary product, supplied free of charge to 'qualified' users – typically organizations with Department of Defense contracts. The point of contact for making a formal request is: Mr. John Lockett, Army Research Labs, Aberdeen, MD.

Other Comparisons

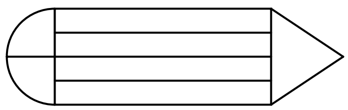
Comparisons can be made on at least four levels: graphic modeling of activities, common features, unique features, and the user interface, for both input and output.

Graphic Modeling

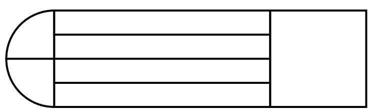
SAINT provided no assistance in developing the network diagrams. A symbol set (Figure 1) was specified for representing task networks, but the diagrams had to be done manually. SADT tools later made this easier to do, but the translation into SAINT was neither direct nor automatic. Micro Saint and IMPRINT both provide facilities for creating the network diagrams.



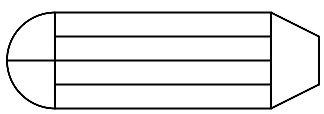
Deterministic Node: Take ALL Branches Out



Probabilistic Node: Take One and Only One Branch, Based Upon the Relative Frequency Specified



Conditional Node: Take ALL Branches for Which the Stipulated Condition Is Met



Conditional Node: Take 1st Branch for Which the Stipulated Condition Is Met

Figure 1. *SAINT Symbols*.

The first node in a network starts the task sequencing; subsequent nodes are ‘triggered’ upon completion of one or more preceding task(s). Directed arrows point from one task node to the task(s) node(s) which follow, and the specified branching logic is applied to determine which path(s) are to be taken. Information packets follow along those paths (like tokens in Petri Nets). These packets are a vehicle for transmission of local information from one task to another (e.g., the level of stress, the value of a control setting, or other definitions for a variable’s value). Subsequent tasks can then examine the values of variables passed in a packet. The value can then influence either the time taken by that task, some variable manipulated in the performance of the task, or some condition tested to determine branching out of the task. Micro Saint retained this capability. It provides a very powerful modeling tool.

When any one task completes, one or more of the subsequent tasks may be released for execution, depending on the precedence requirements or release conditions specified for each task. At some point, a terminating node is reached which ends the simulation and initiates the generation of summary statistics for a series of runs / iterations.

When a continuous system’s dynamics were modeled, SAINT would also generate a ‘strip chart’ recording that showed the level (value) of each continuous variable over time, from the start of the simulation to its termination: the time trajectory for each state variable of interest.

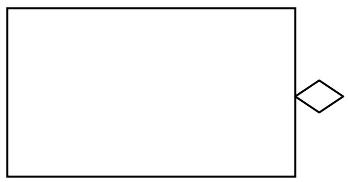
The semi-circular left side of all blocks had an upper and lower half specifying what precedence constraints had to be satisfied (what number of preceding tasks had to be first completed before this task was started). In the upper half, one specified how many of the incoming ‘signals’ had to be present before the current task could be ‘released for execution the first time it was performed. The lower half specified how many had to be present before subsequent releases.

Micro Saint did not distinguish between first and subsequent task execution precedence constraints. Task release could instead be specified on the basis of a specified variable, which if ‘true’ when tested, the task would be released. IMPRINT allows this same representation.

Also, Micro Saint did not use alternate shapes to represent branching alternates. Instead, a dialogue box is presented for the user to select what type of branching is desired. Conditional ‘take first’ branching is not one of the options however. Prioritized branching can be accomplished through setting and testing variables instead. This scheme simplifies the diagram, leaving details to be specified in terms of data inputs. IMPRINT does the same.

Micro Saint IV and IMPRINT use two different shapes to model functions and tasks. Tasks are always a decomposition (more detailed representation) of functions. Figure 2 shows the older Micro Saint symbology, also used in IMPRINT. Micro Saint Sharp is similar but slightly different. A Node may be either a task or a network of tasks. That network can be either a set of tasks, set of networks of tasks or some combination: a powerful hierarchical approach to modeling complex systems.

Function and Decision Block



Task and Decision Block

Figure 2. *Micro Saint Symbology.*

By clicking on either the function or the task blocks, the user will bring up the dialogue box that permits entering data associated with the selected function or task. Correspondingly, by clicking on the diamond to the right of each box, the user calls up the dialogue box for specifying the desired branching logic. While this notation simplifies the diagram, it effectively hides the nature of the branching logic.

SAINT did not preclude hierarchical decomposition of task networks, but neither did it facilitate that kind of modeling. Micro Saint distinguishes between upper level functions and the lower level tasks that then support or implement those functions: a network of tasks at one level can appear as a single task at another level. IMPRINT does this too, but in a more limited fashion (a single function layer and a single task layer). However, when using the Advanced IMPRINT workload assessment technique, only a single task layer is allowed, but parallel paths are permitted. The CART-based mode adds yet another consideration: goals drive which functions may be activated at any one time. Several functions may be ongoing at one time, along with their associated tasks. This allows better representation of mission scenarios, but it also is a more complicated form of modeling and typically requires attention to detail and more time in debugging the implementation.

Common Features

SAINT, Micro Saint, and IMPRINT all provide modelers with a wide variety of statistical distributions for representing the duration of tasks or other activities. All three techniques also offer users flexible ways to adjust the parameters of those distributions to reflect the impact of a wide variety of moderators or stressors that change the character of behavior, either in terms of the duration of the task or in the branching that occurs when a task is completed.

Unique Features

System dynamics portray the states of a system (e.g. airplane) continuously over time. The first example used in SAINT was aerial refueling of a B-52 by a KC-135. What was critical was representing the vertical and longitudinal separation between the two airplanes as the pilot changed yoke and throttle settings. Those discrete tasks changed acceleration characteristics, which affected the speed and vertical velocity of the bomber with respect to the tanker, which in turn altered the position of the bomber relative to the tanker (their separation).

While SAINT provided symbols for modeling discrete activities, continuous processes can also be diagrammed, but SAINT assumed users would use either analogue computer techniques (integrator symbols, logic gates, etc.) or the 'flow rate and level' symbology used by Forester (1961). For simulation, the differential equations portraying system dynamics are expressed as difference equations for integration of rates to get states. Neither Micro Saint nor IMPRINT provide this capability directly.

IMPRINT on the other hand has a built-in ability to reflect the effects of task accuracy (or, conversely, failure) on performance. This feature appears intuitive on the surface, but users should carefully examine this function to be sure what they think it does is what is actually happening. While the explanations provided seem clear, the user would do well to empirically test a simple model to be sure what they expect will happen actually occurs. Otherwise, they need to reinterpret how this function really works!

The User Interface

SAINT required punched card input, and a single typing mistake meant punching a new card and perhaps rerunning the program. Pre-defined outputs were generated on pre-punched computer paper, not regular 8 ½ x 11 inch sheets. The horizontal format provided more space for printing output, but storage of massive output listings was then awkward.

Micro Saint was designed to operate interactively using a PC's display screen. Modifications to input could be made more easily, and turn-around for modeling improved greatly. Results could be displayed before printing, so less paper was wasted, and printers started using more conventional sizes of paper. Considerable flexibility is provided to the user in generating output products, including animation.

IMPRINT is more restrictive and directive than Micro Saint and has its own unique user interface, but that also permits more rapid model development with that standardization. The three different IMPRINT modes do have differences in both the input and output interfaces available to users. Each is tailored to the specific mode being exercised, and animation is not provide except in its simplest form: seeing which task(s) get executed. However, this can be quite useful in debugging model implementation.

IMPRINT addresses two important kinds of application: a) system modeling of an individual performing a specific mission, and b) a series of ongoing engagements where the break and fix rates of malfunctioning equipment determine the ability to sustain combat operations. Each use of IMPRINT has its own special characteristics, data input requirements, and output reports.

Conclusions

While SAINT started with the objective of providing engineers with tools that would permit system modeling that also treated human factors, Micro Saint made this approach to simulation and analysis more practical for both engineers and human factors specialists, and IMPRINT tailored the Micro Saint technology to specific needs of the human factors engineer in systems acquisition programs. Micro Saint Sharp has provided a significant advance over the older versions of SAINT but without incorporating its continuous / combined modeling capabilities. While IMPRINT is a bit more restrictive than Micro Saint, it handles workload well, has been extended to treat the degrading effects of a variety of stressors, and addresses some of the impacts training and the lack of practice can have on performance. Version 8 of IMPRINT, now under development, will incorporate some of the features found in Micro Saint Sharp and should therefore be an even more powerful and flexible tool for modeling and analyzing human performance in the context of mission simulation.

References

Anonymous (2003a), *IMPRINT: Improved Performance Research Integration Tool, Analysis Guide, Version 7 DRAFT,* US Army Research Laboratory, Aberdeen, MD, October.

Anonymous (2003b), *IMPRINT: Improved Performance Research Integration Tool, Users Guide, Version 7 DRAFT*, US Army Research Laboratory, Aberdeen, MD, October.

Bloechle, Wendy K. and Daniel Schunk (2003), "Micro Saint Sharp Simulation", *Proceedings of the Winter Simulation Conference*, New Orleans, LA, 7-10 December.

Booher, Harold R., editor (1990), *MANPRINT: An Approach to Systems Integration*, Van Nostrand Reinhold, New York.

Brett, Bryan E., Jeffrey A. Doyal, David A. Malek, Edward A. Martin, David G. Hoagland, and Martin N. Anesgart (2002), *The Combat Automation Requirements Testbed (CART): Task 5 Interim Report – Modeling a Strike Fighter Pilot Conducting a Time Critical Target Mission*, AFRL-HE-WP-TR-2002-0018, Crew Systems Interface Division, Wright-Patterson Air Force Base, OH.

Cellier, F. E., editor (1982), *Progress in Modeling and Simulation*, Academic Press, New York.

Chubb, Gerald P. (1971), *F-106A Nuclear Vulnerability Analysis, Vol. VIII: Crew Effectiveness,* AFSWC-TR-70-7, Air Force Weapons Laboratory, Kirtland AFB, NM.

Chubb, Gerald P. (1989), "SAINT Performance Assessment Model of a SAM System," in Stephen A. Murtaugh and Sally VanNostrand, editors, *Proceedings of MORIMOC II: MORS Symposium on Human Behavior and Performance as Essential Ingredients in Realistic Combat Modeling*, Center for Naval Analyses, Alexandria, VA. 22-24 February, Volume I, pp. 199-219.

Desrochers, Alan A. and Robert Y. Al-Jaar (1995), *Applications of Petri Nets in Manufacturing Systems: Modeling, Control, and Performance Analysis*, IEEE Press, New York.

Forrester , Jay Wright (1961), *Industrial Dynamics*, MIT Press, Cambridge, MA.

Hoyland, Constance M., Debbie Ganote, and Gerald P. Chubb (1988), "C-SAINT: A Simulation Tool Customized for Workload and Information Flow Analysis," Proceedings of the IEEE National Aerospace and Electronics Conference (NAECON). Dayton, OH, 21-25 May, pp. 823-830.

Kelley, Troy D. and David R. Scribner (2003), *Developing a Predictive Model of Dual Task*

Performance, ARL-MR-0056, US Army Research Laboratory, Aberdeen, MD, September.

Law, Averill M. and W. David Kelton (2000), third edition, *Simulation Modeling & Analysis*, McGraw-Hill, New York.

Laughery, "Network modeling on microcomputers," *Simulation*, January, pp. 10-16.

Marca, David A., Clement L. McGowan (1988), *SADT: Structured Analysis and Design Technique*, McGraw-Hill, New York.

McCracken, J. H. and T. B. Aldrich (1984), *Analyses of Selected LHX Mission Functions: Implications for Operator Workload and System Automation Goals*, Tech Note ASI479-024-84, Army Research Institute, Aviation Research and Development Activity: Ft. Rucker, AL.

Moder, Joseph J., Cecil R. Phillips, and Edward W. Davis (1983), *Project Management with CPM, PERT, and Precedence Diagramming*, Van Nostrand Reinhold, New York.

North, R. A. and V. Riley (1989), "W/INDEX: A Predictive Model of Operator Workload", In G. R. McMillan, editor, *Applications of Human Performance Models to System Design*, Plenum Press, New York.

Pritsker, A. Alan B. and W. W. Happ (1966), "GERT: Graphical Evaluation and Review Technique – Part I. Fundamentals, *Industrial Engineering*, Vol. XVII, June, pp. 267-274.

Pritsker, A. Alan B. and G. E. Whitehouse (1966), "GERT: Graphical Evaluation and Review Technique – Part II. Probabilistic and Industrial Engineering Applications, *Industrial Engineering*, Vol. XVII, June, pp. 293-301.

Pritsker, A. Alan B. and Nicholas R. Hurst (1973), "GASP IV: A Combined Continuous-discrete FORTRAN-based Simulation Language", *Simulation*, Vol. 21, September, pp. 65-70.

Prtisker, A. Alan B. (1995), *Introduction to Simulation and SLAM II*, John Wliey & Sons, New York.

Seifert, Deborah J. and Gerald P. Chubb (1978), "SAINT: A Combined Simulation Language for Modeling Large, Complex Systems," *Proceedings of the SIGSIM 78 Conference*, Canberra, Australia. 4-8 September (also identified as AFAMRL-TR-78-48, Wright-Patterson AFB, OH).

Siegel, Arthur I. and J. Jay Wolf (1969), *Man-Machine Simulation Models; Psychosocial and Performance Interaction*, Wiley-Interscience, New York.

Wickens, C. D. (1984) "Processing Resources in Attention", In R. Parasuraman, J. Beaty, R. Davies, editors, *Varieties of Attention*, Wiley, New York, pp. 63-101.

Wickens, C. D. and Y-Y Yeh (1986), "Multiple Resource Model of Workload Prediction and Assessment" in *IEEE Conference on Systems, Man, and Cybernetics* (Atlanta, GA.)

Yourdan, Edward and Larry L. Constantine (1979), *Structured Design: Fundamentals of a Discipline of Computer Program and Systems Design*, Prentice Hall, Englewood Cliffs, N.J.

IS PILOTS' VISUAL SCANNING ADEQUATE TO AVOID MID-AIR COLLISIONS?

Kurt Colvin[1], **Rahul Dodhia**[2], **R. Key Dismukes**[3]

The "See and Avoid" concept is crucial to visual meteorological condition (VMC) operations. The FAA and other organizations prescribe a specific systematic out the window (OTW) visual scanning pattern to avoid traffic conflicts, however little research has been published on what scanning patterns pilots actually use and how effective their scanning is. In our study, commercial pilots flew VFR scenarios in a general aviation flight training device (GAFTD) equipped with head and eye tracking equipment. We developed new algorithms to analyze the effectiveness and patterns of visual scanning. The scanning patterns used by the participant pilots did not resemble the prescribed patterns.

Introduction

The "see and avoid" concept remains the primary defense against mid-air collisions in VMC. Although airliners and many corporate aircraft are now equipped with Traffic Collision Avoidance Systems (TCAS) that alert crews to the presence of conflicts with aircraft with an operating transponder, these systems are intended to supplement rather than replace "see and avoid". Further, most light aircraft are not equipped with TCAS because of the expense.

The FAA and other organizations recommend a systematic visual search scan for traffic in which the pilot fixates at a location for at least one second, then shifts gaze no more than 10 degrees in order to sequentially scan the entire the visual field outside the window. Pilots are advised to look inside the cockpit no more than 4-5 seconds for every 16 seconds spent scanning the outside world (FAA, 1998a; AOPA, 2001). Although all pilots are exposed to this concept, they do not receive systematic or extensive training in how to execute and maintain it over long periods in coordination with other cockpit tasks. Humans are notoriously poor at maintaining vigilance in searching for targets or monitoring for events that rarely occur (Baker, 1960; Smith, 1969). Further, the type of scan traditionally recommended requires considerable cognitive effort, competing with other cockpit task demands, It is not known to what extent, if any, pilots may be able to learn to scan automatically, which would reduce cognitive effort. Thus it would be highly desirable to learn what scanning patterns pilots actually use and how effective those patterns are. To date little research has been reported to this end.

Previous studies that have used eye tracking have focused primarily on monitoring of cockpit systems and displays; however, some of these studies included measures of percentage of time spent looking outside the cockpit and found that this percentage is substantially less than the FAA's recommendation (e.g., Wickens, et. al, 2000; Mumaw, Sarter & Wickens, 2001; Anders, 2001). Howell (1957) conducted an actual flight study in which pilots encountered conflicts arranged by the experimenter with other aircraft. Of the 128 conflict trials, nine (7%) ended without the participant pilot detecting the conflict (the experimenter arranged for the conflict to terminate before safety was compromised). On successful trials the average detection distance varied from 3.4 to 5.4 miles, and performance was not affected by whether the pilots were informed that they would encounter traffic.

Sophisticated navigation equipment and "glass cockpit" displays are rapidly coming into use in light general aviation aircraft. This equipment is generally more complicated than traditional systems, and pilots are vulnerable to becoming preoccupied with using this equipment and remain head-down for prolonged periods. This development may require greater emphasis in training on maintaining effective visual scanning, however development of better training requires better understanding of how scanning is accomplished and of the nature of vulnerability to lapses in scanning.

This paper provides an update on our continuing project to investigate pilots' visual scanning behaviors, using eye tracking as pilots fly in a GAFTD (Colvin et al., 2003). Our goal is to determine what patterns pilots use, differences among pilots, the effects of various conditions on scanning, and the adequacy of scanning to avoid conflicts with other aircraft. This paper focuses on the adequacy of scanning, reports considerable differences among pilots, and provides preliminary data on patterns of scanning. We are developing new ways to measure and evaluate these functions, and report here a measure of the fraction of time the outside world was adequately searched. Determining scanning patterns turns out to be difficult because these patterns vary enormously moment to moment. We found large differences among pilots in adequacy of scanning, with most pilots failing much of the time to scan

[1]California Polytechnic Sate University, San Luis Opispo, California; [2]San Jose State Foundation/NASA Ames Research Center, Moffett Field, California; [3]NASA Ames Research Center, Moffett Field, California

frequently enough in the lateral dimension to detect conflicting aircraft.

Methods

Participants

Twelve pilots were recruited and paid to participate in the experiment. All possessed at least a current FAA instrument rating with appropriate airplane ratings and had 20/20 visual acuity or were corrected to that value. The median of their total flight hours was 1400, and the median number of years flying was 15.

Apparatus

Eye tracking data were collected using the ISCAN, Inc. Line Of Sight (LOS) system. This equipment consists of a headband fitted with a camera to determine the eye position, a magnetic sensor to determine head orientation and a computer that performs the computations necessary to determine where the pilot is looking in the cockpit. To facilitate analysis, the cockpit was divided into six two-dimensional planes, referred to as the areas of interest (AOIs). Four of these AOIs were the GAFTD's windscreens displaying the "outside" visual world, and the other two AOIs were the instrument and engine indicator panels. The LOS system calculates the plane to which gaze is directed, the location of gaze within the plane (X and Y coordinates), and pupil diameter of the eye. These parameters are sampled at a rate of 60 Hz.

An AST Hawk 201 FAA-approved flight-training device was used to simulate a high performance, complex single-engine piston aircraft. The four cockpit windows have 17" CRTs that depict the scene outside the window, including terrain, sky, and traffic, as programmed.

Procedure

Participants were given written instructions that emphasized that they were to perform all tasks, including scanning for traffic, just as they would in actual flight. They then flew a scripted 45-minute training session to familiarize them with the GAFTD, after which they were calibrated on the eye-tracking apparatus.

Participants then flew the experimental scenario, a 45-minute VFR cross-country flight in which they navigated by reference to VORs on a flight plan without interacting with ATC. After reaching cruise altitude, participants encountered in sequence a low workload period (LWL1 – 3 minutes), a high workload period created by moderate turbulence in the vicinity of high terrain (TURB – 3 minutes), a second low workload period (LWL2 – 3 minutes), a traffic period (TRAFFIC – 14 minutes), and a final low workload period (LWL3 – 3 minutes). During the traffic sequence, aircraft appeared for periods ranging from 43 to 75 seconds at various crossing angles. Nine aircraft appeared, one at a time, with 30 seconds between aircraft. These aircraft were traveling level at either 500 feet or 1000 feet above or below the participants' aircraft, however it was not initially obvious that the aircraft were not on a collision course.

Results

Eye fixations were extracted and saccades were eliminated from the raw data by the absolute deviation method (Salvucci & Goldberg, 2000). A clustered sequence of data points is counted as a fixation if the absolute deviation of the cluster is less than one degree of visual angle and the duration of the sequence is greater than 100msec. This analysis results in four parameters for each fixation—area of interest (plane), mean horizontal and vertical location within the plane and fixation duration—that are used by our algorithms for calculating spatial and temporal patterns of eye movements.

We report here four measures of visual scanning:

1) Percent time that fixations are directed toward the cockpit windscreens (denoted as percent time OTW.

2) Distribution of fixations over AOIs.

3) Fraction of time each part of the outside world is searched safely. This is calculated by first determining a "grace period": the time from the first moment that a pilot would be likely to detect an aircraft if fixating gaze at or near the aircraft's position to the time of collision, minus the time required to execute an evasive maneuver. If, after having fixated a point outside the windscreen, the pilot returns gaze to that point (within 2.5 degrees) before the grace period is over, that area of space has been searched frequently enough to avoid collision. If gaze does not return before the end of the grace period, that area of space is considered unsafe until re-fixated. The fraction of time each area of space is searched safely is calculated, using specific assumptions about parameters. We selected six miles as the average distance at which pilots could reliably detect another aircraft in most daytime meteorological conditions, and used a combined

closure rate of 385 knots, representing a conflict that might occur between light aircraft and transport aircraft below 10,000 feet. (The closure rate only varies by a factor of 15% for collision angles between zero (head-on) and 40 degrees.) Little published data is available on the range at which pilots can detect aircraft. Harris' data (1973) suggest that pilots would have about an 86% chance of detecting a DC-3 if fixating the target at six miles, and Andrews (1977) data also suggest that six miles is a reasonable approximation. We allowed 15 seconds as the average time a pilot would require recognize an aircraft, determine that if it is on a collision course, and complete an avoidance maneuver (FAA 1998a). Other assumptions about detection range, rate of closure, and response time can easily be substituted in our algorithm.

4) A transition matrix depicting the relative proportion of transitions from one AOI to each of the other AOIs.

Results

Figure 1 shows that on average participants spent just under one third of their time looking outside the cockpit, except during the traffic period, in which looking outside increased to 51%. When participants detected traffic they monitored the path of the observed aircraft, increasing the total percentage of time looking outside. However when traffic ceased, the percentage of time looking outside again dropped. During the time not spent looking outside, fixations were predominantly directed to the instrument panel (data not shown). The standard deviation bars on the figure reveal large variation among the 12 participants.

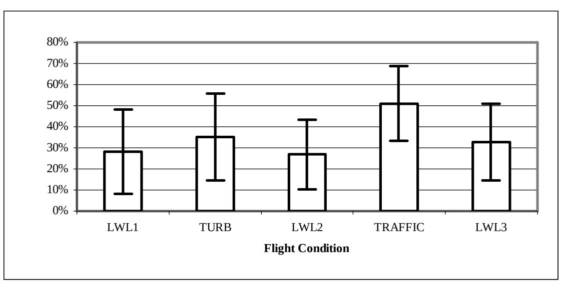
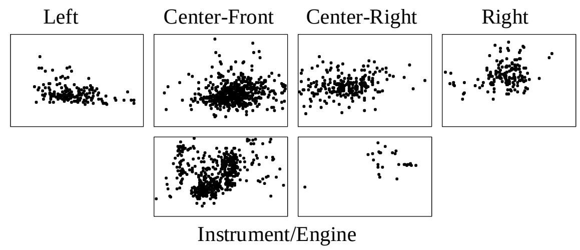
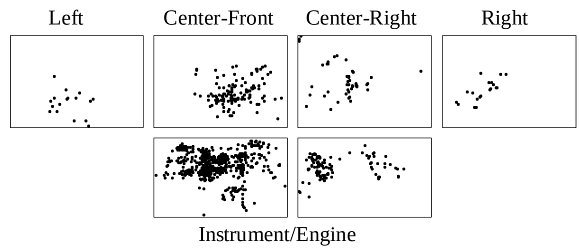


Figure 1. ***Percent time gaze directed out the windscreen***

Figure 2 is a scatter plot of fixations over the six AOIs during low workload periods for two participants, one that spent the great majority of his time gazing at the instrument panel, and another that distributed his gaze primarily outside. Both pilots directed gaze more often to the center-front windscreen AOI than to the other three windscreens combined. Outside fixations tended to line up with the horizon, with relatively few fixations being directed to either the top or the bottom of the windscreen. Also, fixations tended to cluster more toward the center of the windscreen than to either side.



(a) – Best performer



(b) – Worst Performer

Figure 2. *Total fixations of two participants during combined low workload periods (IP: instrument panel; EP: engine indicator panel).*

Figure 3 depicts the average rate of fixation on the six AOIs during low workload periods. The standard deviation bars reveal large variation among participants, however this variation is driven more by the relative distribution of gaze between the instrument panel and outside the windscreen than by variation in distribution of gaze across the windscreens (data not shown). On average participants fixed the instrument panel far more frequently than the windscreens, and they fixated the center-front windscreen far more frequently than the other three windscreens.

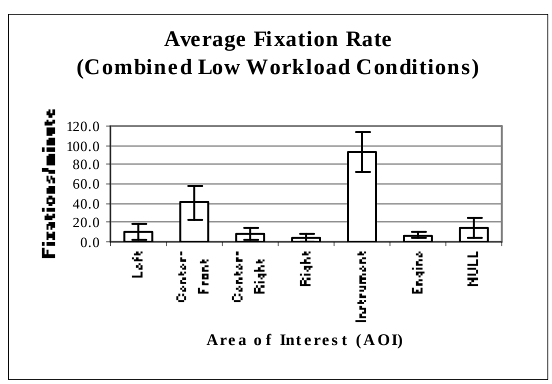


Figure 3. ***Average Fixation Rate***

Our metric of the fraction of time the outside world was safely searched reveals substantial differences in scanning of the four windscreens and in scanning within each of the windscreens (Figure 4). On average, pilots' scanning of the center-front windscreen was adequate most, though not all, of the time, and scanning of the left and right sides of this windscreen was adequate less than 50% of the time. Scanning of each of the windscreens tended to favor the center of the display over the edges. The asymmetry is largest vertically, however, analysis of collision geometries for rates of climb and descent typical of civil aircraft reveals that pilots need search only about three degrees above and below the horizon to avoid collisions (Fries, 2004).

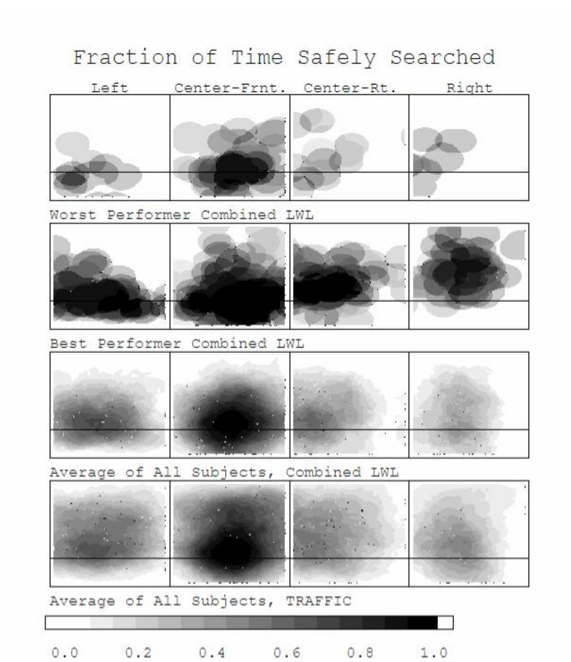


Figure 4. *Fraction of time safely searched.*

Scanning the off-center windscreens was much less adequate: Fraction of time adequately scanned ranged from around 0.5 for the left windscreen, to around 0.4 for the center-right, to around 0.3 for the right (values at the center of each windscreen). Scanning was even less adequate near the left and right edges of each windscreen. However, participants varied greatly in adequacy of scanning. The best performer scanned all four windscreens adequately the great majority of the time, though he also scanned the left and right sides of the windscreens less often than the centers. The worst performer would have had little chance of detecting a traffic conflict except for a head-on collision course.

The transition matrix shows a strong tendency for gaze to return to the center windscreen from whatever other AOI was previously fixated (Figure 5). When gaze exited the center windscreen it predominantly went to the instrument panel and vice versa. (The instrument panel and engine indicator panel were combined for this analysis. Gaze was directed to the engine indicator panel far less than to the instrument panel). One-step transitions (moving from one windscreen to another immediately adjacent) predominated over two-step and three-step transitions (jumping over adjacent windscreens), and this weighting remained even after correction for the fewer number of two-step and three-step transitions possible (correction data not shown).

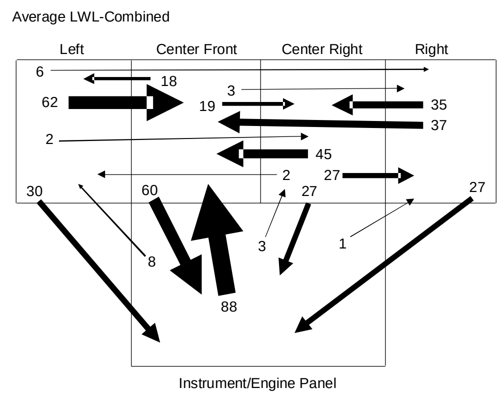


Figure 5. *Transition matrix for all participants, combined low workload periods*

Discussion

Although not one of the largest causes of accidents, mid-air collisions persist in general aviation, around 15 per year (FAA, 1998b), and are usually fatal. Increasing traffic density, as may occur now that the FAA has published the light sport aircraft rule, will increase this threat. The "see and avoid" concept is the primary defense against mid-air collision for aircraft operating under VFR. Does the relatively low (though still unacceptable) number of mid-air collisions indicate that the see and avoid concept generally works well, or merely that mid-air collisions are fairly unlikely because of the "big sky" in uncongested areas? To what extent do pilots use the visual scanning technique recommended by the FAA and other organizations, and to what extent is it even practical to use this technique, especially in coordination with other cockpit tasks? If pilots use other scanning patterns, how effective are these?

Few research data exist to answer these questions, however previous eye-tracking studies in flight simulators/training devices suggest that pilots look outside less often than recommended (Wickens, et. al, 2000; Mumaw, Sarter & Wickens, 2001; Anders, 2001). Furthermore, Howell's (1957) empirical study of actual airborne conflicts found that pilots did not always detect conflicting traffic.

Our data, consistent with previous studies, reveal that pilots participating in this study, spent more time looking inside the cockpit than outside. This behavior was probably influenced to some degree by the requirement that they follow a VFR flight plan, navigating by VORs. It is also conceivable that pilots did not use the scanning patterns they normally employ in actual flight, perhaps not thinking traffic detection to be important in a simulation. However two facts argue against this possibility: (1) we emphasized in our instructions that participants were to perform all normal flight duties, including watching for traffic, and (2) when participants observed traffic they monitored the course of that traffic.

Our several measures provide converging evidence on the visual scanning performance of the participants. Large differences occurred among the participants: Scanning by the best performer was largely though not completely adequate; scanning by the worst performer was abysmal, and the average left participants vulnerable to not detecting conflicting aircraft quickly enough to avoid a collision much of the time.

Scanning the outside world strongly favored looking straight ahead, with many fixations directed only a few degrees to either side. We suspect that many of these fixations represent not searching for traffic but rather the default position for gaze, centered along the central axis of the pilot, the aircraft, and the direction of travel. Gazing mainly straight ahead, coupled with peripheral vision, allows pilots to maintain control of the aircraft.

All participants did scan all windscreens to some degree; averaged data show the distribution of these fixations to be centered just above the horizon and to relatively neglect the edges of the windscreen. The neglect of the upper part of the windscreens is not problematic: typical rates of descent for civil aircraft would not allow collision for vertical angles much more than about three degrees above the horizon (Fries, 2004). However, neglect of the left and right sides of the windscreen is more problematic. We suspect it occurs because the windscreen provides a frame that guides gaze toward its center. This left and right neglect, coupled with other data from this study, suggests that participants did not systematically scan small segments of the outside world sequentially.

We are still working to analyze the patterns of visual scanning. The transition matrix shows that participants were not consistently following a systematic left-to-right or right-to-left scan, however the matrix does not eliminate other types of systematic scanning, such as a pattern in which the participant would look from the center to another windscreen, back to the center, then on to the next windscreen, back to the center, and then on the last windscreen. However, even if participants followed this pattern some of the time, they clearly were not following it most of the time, because of the relative neglect of the outer windscreens.

Conceivably the transitions among AOIs are random, driven only by the relative probabilities of each type of transition. However, we have conducted a preliminary Markhov analysis that indicates that the probability of transition from one AOI to another is partially influenced by which AOI was previously fixated. This suggests some sort of patterns longer than single transitions do occur. We are currently analyzing the sequences of transitions among AOIs, and so far have found that many different patterns of different chain lengths occur. The data are very noisy, indicating that participants are not following a single or even a few scan patterns.

We do not find it surprising that participants did not use the FAA recommended scan pattern. The recommended scan pattern requires considerable cognitive effort. In the absence of frequent traffic, whose detection would provide a positive feedback loop, scanning becomes a vigilance task, and humans are well known to be poor at maintaining vigilance beyond short periods. Further, effortful visual scanning must compete with other cockpit tasks for limited cognitive resources.

Some caution is required in interpreting our results. Conceivably our sample of 12 pilots does not well represent general aviation pilots, although their flight experience probably exceeds the average. Also it is conceivable that our participants did not scan as well in the GAFTD or in this scenario as they normally do in actual flight. However, if these participants' performance is indeed representative, our data suggest that the relatively low (though unacceptable) rate of mid-air collisions in general aviation aircraft not equipped with TCAS is as much a function of the "big sky" as it is of effective visual scanning.

Lest this analysis sound too pessimistic, we raise the possibility that pilots may, through practice, may develop scanning techniques that can be executed largely automatically, reducing the demand for limited cognitive resources and perhaps making it possible to maintain the scan with little overt attention. Conceivably the more effective scanners in our study had developed such techniques on their own—we are currently investigating that possibility.

Acknowledgments

We thank Kim Jobe for help in preparing this manuscript and Sean Belcher for helping design the flight scenario to collect data. This research was supported by the Human Measurement and Performance element of NASA's Aviation Systems Program.

References

Anders, G. (2001). Pilot's attention allocation during approach and landing – eye- and head-tracking research in an A330 full flight simulator. Conference Proceedings on CD-ROM of the 11th International Symposium on Aviation Psychology, Columbus, OH: The Ohio State University Press.

Andrews, J.W. (1977). Air-to-air visual acquisition performance with pilot warning instruments (PWI). Federal Aviation Administration Technical Report (FAA-RD-77-30).

AOPA (2001). Collision Avoidance: Strategies and Tactics. Safety Advisor, Operations and Proficiency No. 4. Published by Aircraft Owners and Pilot's Association, Frederick, MD.

Baker, C. H. (1960). Toward a Theory of Vigilance. Visual Search Techniques. Symposium Conducted at The Spring Meeting, 1959. Committee on Vision. Division of Behavioral Sciences National Research Council. National Academy of Sciences. Washington, D.C. 1959.

Colvin, K. W., Dodhia, R. M., Belcher, S. A., & Dismukes, R. K. (2003). Scanning for visual traffic: An eye tracking study. In *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 255-260). Dayton, OH: The Wright State University.

Fries, Ian Blair (2004). Scanning for traffic. AOPA Pilot. Vol. 47, No10. Fredrick, Maryland.

Federal Aviation Administration (1998a). Scanning for Other Aircraft. Aeronautical Information Manual, 8-1-6-c. Oklahoma City, OK: Author.
Federal Aviation Administration (1998b). Aviation systems indicators: 1998 annual report. Washington, DC: Author.

Harris, J.L. (1973). Visual aspects of air collision. *Visual Search*, National Academy of Sciences, Washington, D.C.

Howell, W.D. (1957). Determination of Daytime Conspicuity of Transport Aircraft. Technical Development Report No. 304. Civil Aeronautics Administration Technical Development Center. Indianapolis, Indiana.

Mumaw, R. J, Sarter, N. B, & Wickens, C. D. (2001). Analysis of pilots' monitoring and performance on an automated flight deck. Conference Proceedings on CD-ROM of the 11th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Salvucci, D.D. & Goldberg, J.H. (2000). Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the Eye Tracking Research and Applications Symposium, (pp. 71-78). New York: ACM Press.

Smith, R.L. and L.F. Lucacini (1969). Vigilance research: its application to industrial problems. Human Factors, Vol. 11, No. 2. The Johns Hopkins University Press.

Wickens, C.D., Xu, X., Helleberg, J. R., Carbonari, R. & Marsh R. (2000). The allocation of visual attention for aircraft traffic monitoring and avoidance: Baseline measures and implications for freeflight. (Technical Report ARL-00-2/FAA-00-2). Savoy, IL: University of Illinois, Aviation Research Lab.

STUDY ON THE INTEGRATION OF HUMAN PERFORMANCE AND ACCIDENT RISK ASSESSMENT MODELS: AIR-MIDAS & TOPAZ

Kevin Corker
San Jose State University
Henk A.P. Blom & Sybert H. Stroeve
National Aerospace Laboratory NLR
Amsterdam, The Netherlands

A computational model of human performance (Air Man-machine Integration Design and Analysis System, Air MIDAS) and an accident risk assessment methodology (Traffic Organization and Perturbation AnalyZer, TOPAZ) were integrated in order to learn about the similarities and differences of their models, to demonstrate the feasibility of such integration, and the integration impact on accident risk assessment.

Introduction

In the analysis and design of advanced operations in complex, dynamic, human-machine systems, accident risk assessment is a critical component of effective system engineering. Probability Risk Assessment (PRA) techniques typically model such complex system by assigning conditional probabilities of the success, or failure for system operations into fault and event trees (e.g. Kumamoto and Henley, 1996). Subsequently, an assessment of risk is undertaken by evaluating the combined effects of the conditional probabilities in these fault and event trees. The role and contribution of the human operator has proven to be a significant element to both accident risk (Hollnagel, 1993), and to system safety and effectiveness (Dekker 2001).

The development of models that represent the contribution of the human operator to risk has been explored for some 30 years (Swain & Guttman 1983). The function of the human operator was either assigned a probability of success or failure, as would be provided for any system component, and the "integration" was the inclusion of those probabilities in the overall system success failure assessment. A serious limitation of fault and event tree based PRA is its inability to evaluate the effects of concurrent and dynamic behavior on accident risk. The remedy is to exploit stochastic dynamical modeling and Monte Carlo simulation of the concurrent and dynamic processes for accident risk assessment (e.g. Labeau et al., 2000) and to include explicit representation of human performance in individual and on teams (e.g.,. Cacciabue, 1998 or Corker, 2000)

In order to apply this approach to air traffic management, multiple human operators and their interactions with each other and with aircraft and ground systems have to be modeled and simulated.
Both with the human performance model Air-MIDAS (Corker, 2000) and with the accident risk assessment methodology TOPAZ (Blom et al., 2001, 2003; Stroeve et al., 2003), significant and complementary headway has been made. We report here on the integration of Air-MIDAS and TOPAZ in aviation safety assessment.

The objective of this integration is to combine the significant advances established in individual human performance representation and human performance factors (human factors in general and human cognitive behavior in particular) through large-scale simulations for accident risk assessment. As an objective test for the success of this integration we hypothesize that this combination allows Air-MIDAS to provide simulation results for individual human operators which improves the accident risk assessment.

The aviation community continues to be concerned with accident risk and runway operations and several technologies have been under development to mitigate this risk. Given the relevance of these operations to both safety risk and human performance, an integrated simulation of the baseline conditions for runway incursion avoidance was undertaken by Air-MIDAS and TOPAZ simulation toolset TAXIR for this operation.

Integration of human modeling approaches

Because of the complementary objectives and separate developments of Air-MIDAS and TOPAZ their human performance modeling approaches show similarities and differences. Their potentially complimentary functions form the reason why this integration is so useful and challenging at the same time. In the course of the integration study the complementary human performance modeling details of both approaches have become clear. A short explanation of this is given next, including an overview in Table 1.

Table 1 *Human performance modeling in Air-MIDAS and TOPAZ*

		Air-MIDAS	TOPAZ
A	Management modes	Max-load or Even-load	None
	Control Modes	Matching with Rasmussen's SRK (Skill, Rule, Knowledge)	Matching with Hollnagel's tactical and opportunistic control modes
	Switching between modes	Fixed thresholds	Thresholds with hysteresis
B	Task Scheduling	Goal oriented subtask scheduling	Priority rules for aggregated tasks
	Resources model	Multiple: Visual, Auditory, Cognitive, Psychomotor	Aggregation on the basis of time-critical tasks/resources combinations
	Memory model	Procedural (with decay) Declarative (with decay) Knowledge (no decay)	Aggregated (no decay)
C	SA model	SA of one human only	Multi Agent SA and interactions
D	Human error	Is result of detailed modelling	Amalberti's error recovery model is added
E	Behaviour of Non-human entities	Nominal	Nominal & Non-Nominal
F	Specification language	Air-MIDAS specific, based on LISP	Dynamically Coloured Petri Nets (DCPN)

Integration of these approaches ensures that the simulation scenario under examination is jointly represented in the two modeling systems. This allows identification of values for specific parameters of human performance in the TOPAZ simulation model to be supplied by the Air MIDAS simulation. These parameter values are generated in Monte Carlo runs of the human performance model and subsequently supplied as input to improving TOPAZ simulations. In so far as the modeling paradigms allow similar representation, this parameter exchange is straightforward. For example, simulation of pilot reaction time to recognition of an incursion by the taxiing aircraft is represented in both modeling processes, hence reaction time is a straight forward parameter value to exchange.

Application context

The following operational concept for crossing of an active runway is being considered. A simplified representation of the runway configuration is used, as shown in Figure 1. It consists of one runway with a crossing at a length y_3^b from the runway start threshold. The crossing has remotely controlled stopbars on both sides of the runway. The runway is being used for taking off aircraft. The traffic crossing over the runway accounts for traffic between apron(s) and a second runway. The involved human operators include the start-up controller, the ground controller, per runway a runway controller, the departure controller, and the pilots flying and pilots not flying of taking-off aircraft and crossing aircraft.

Communication between controllers and aircraft crews is via standard VHF R/T. Communication between controllers is supported by telephone lines. Monitoring by the controllers can be by direct visual observation and is supported by radar track plots. Monitoring by the aircraft crews is by visual observation and is supported by the VHF R/T party-line effect.

In the runway crossing operation considered, the control over the crossing aircraft is transferred from the ground controller to the controller of the runway to be crossed. If the runway controller is aware that its runway is not used for a take-off, the crew of an aircraft intending to cross is cleared to do so. The pilot not flying of the crossing aircraft acknowledges the clearance and then the pilot flying initiates the runway crossing. As soon as the crossing aircraft has vacated the runway, then the pilot not flying reports this to the controller of that runway. Next the control over the aircraft is transferred from this runway controller to either another runway controller or to the ground controller.

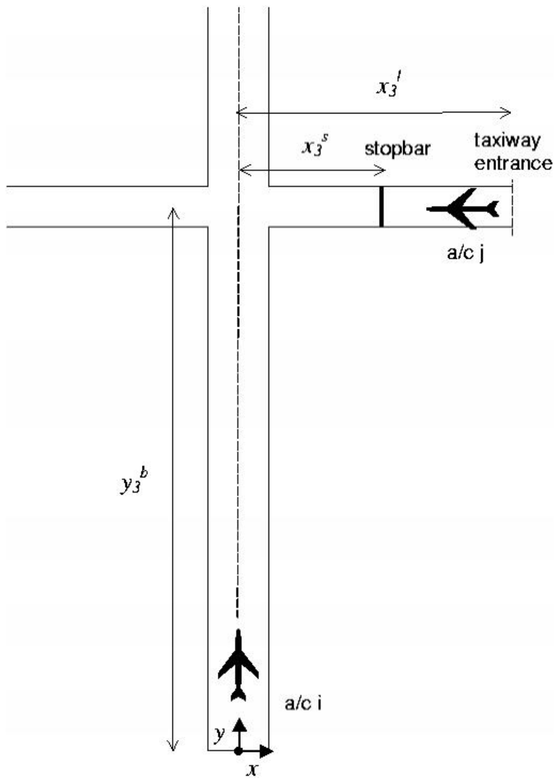


Figure 1: ***Configuration of active runway crossing operation considered. Aircraft i and j respectively take off from a position at the runway start and taxi along a taxiway leading to a runway crossing at a given distance from the runway start.***

Joint model for integrated simulations

The TOPAZ and Air-MIDAS simulation models consider the following human agents: pilots flying for both the taxi aircraft and the taking off aircraft, and the runway controller. The most important elements of these human and other entities are shortly described below.

Initially, the pilot flying (PF) of the taking-off aircraft has the situation awareness (SA) that take-off is allowed and initiates a take-off. During the take-off the PF visually monitors the traffic situation on the runway. During a monitoring action the PF may not observe the crossing aircraft, because of a limited gaze angle or the distance with the crossing aircraft exceeds a viewing threshold, or occasional heads-down time for engine parameter sampling. The monitoring process includes distance dependent error components. Furthermore, the PF monitors the VHF communication channel. The PF of the taking-off aircraft starts a collision avoiding braking action if (s)he observes the crossing aircraft within a critical distance of the runway centre-line or in reaction to a controller clearance, and (s)he decides that braking will stop the aircraft in front of the crossing aircraft.

Initially, the PF of the taxiing aircraft has the SA that either (s)he is taxiing on a regular taxiway, which does not cross a runway or (s)he is taxiing on a taxiway approaching the runway crossing. In the latter case the PF may have the SA that crossing is allowed. Both in the case that the PF has the SA that (s)he is taxiing on a regular taxiway and in the case that the PF is aware that a runway crossing is allowed, the PF proceeds on the runway crossing. During taxiing the PF visually monitors the traffic situation. The characteristics of the monitoring process depend on the SA of the PF concerning the next airport waypoint (either runway crossing or taxiway). After passage of the stopbar the PF may receive a hold clearance by the runway controller. There is a probability that the controller message is not properly understood by the PF. In response to a hold clearance or an observed conflict the PF initiates braking of the aircraft, unless the cockpit of the crossing aircraft is estimated to be already within a critical distance of the runway centre-line.

The runway controller visually monitors the traffic situation on the runway. There is a probability that during monitoring an aircraft is not observed. In response to an alert, the controller directly monitors the traffic situation and the TOPAZ controller model updates the SA. If the controller is aware that the crossing aircraft has passed the stopbar then (s)he specifies a hold clearance to both the crossing and the taking-off aircraft.

Parameters jointly represented

As noted, the representations that the two simulation modeling systems provide are in some ways similar and in others different. Upon examination of the similarities and differences of the models used for the surface operation considered by Air MIDAS and by the TOPZ-TAXIR toolset, a list of model parameters to be affected by the joint runs was identified. These parameters are grouped and provided as follows:

- braking initiation times of pilots flying;
- inter-monitoring time of pilot flying of taxiing aircraft;
- duration of visual observation of pilots flying.

Braking initiation time of PF's This parameter group includes the braking initiation times of pilots flying of taking-off or taxiing aircraft in either tactical of opportunistic mode, when they have become aware of a conflict with the other aircraft. In an overview is provided of the probability density functions (PDF's) and related parameter values for Air-MIDAS and the original and the modified TOPAZ-TAXIR. In all three models equal PDF types and parameter values for the braking initiation times are chosen for the pilots flying of the taking-off and taxiing aircraft, regardless of their cognitive control modes.

It was observed that in comparison to Air-MIDAS, the original TOPAZ-TAXIR has a smaller mean braking initiation time, and a larger tail (probability of more than 5 s initiation time). In order to improve on these aspects, for the modified TOPAZ-TAXIR model the Rayleigh PDF has been selected. The improvements are:

- its shape better fits to the Air-MIDAS data,
- it supports positive values only,
- has a more realistic tail than Gaussian PDF

The parameter value of the Rayleigh PDF has been chosen such that its standard deviation equals the standard deviation of the PDF chosen in Air-MIDAS.

Inter-monitoring time of PF of taxiing aircraft

It is assumed in TOPAZ-TAXIR that the inter-monitoring time of the pilot flying of the taxiing aircraft is independent from the cognitive control mode of the pilot. In the original model this time was represented by an exponential probability density function. Simulations of Air-MIDAS resulted in a data-set of 536 inter-monitoring times of the taxiing pilot flying. These data were well represented by an exponential PDF. Therefore, in the modified model the inter-monitoring times of the taxiing PF are also chosen from an exponential PDF with a mean equal to the estimated mean of the Air-MIDAS data.

Duration of visual observation of PF's

This parameter group includes the visual observation times of pilots flying for the taking-off or taxiing aircraft in either tactical of opportunistic mode. The PDF's of these times in the original model are exponential PDF's with a mean that is smaller in the opportunistic mode than in the tactical mode.
Air-MIDAS simulations provided data on the duration of the tasks:

- 'Monitor Out The Window' for the PF of the taking-off aircraft, and
- 'Decide Action - Decide Take-off Spotted' for the PF of the taxiing aircraft.

These tasks were found to be in good agreement with the visual observation tasks of the pilots flying of the taking-off and taxiing aircraft, respectively. These data were provided for the three control modes used in Air-MIDAS.

Integration Impact on Collision Risk Model

In Table 2 the collision risk results of both versions of TOPAZ-TAXIR are shown for three values of the distance of the runway crossing with respect to the runway start threshold. It follows from these results that the collision risks as evaluated by the modified model are smaller than those evaluated by the original model and that the relative differences in collision risk tend to get larger for larger crossing distances. In all cases, the difference between the results is within a factor two.

Table 2: *Collision risks evaluated by the original and modified TOPAZ-TAXIR models for three crossing distances.*

Crossing distance	**Original Collision Risk** (occurrence per take-off)	**Modified Collision Risk** (occurrence per take-off)
500 m	$1.3\ 10^{-8}$	$1.2\ 10^{-8}$
1000 m	$1.1\ 10^{-8}$	$7.1\ 10^{-9}$
2000 m	$8.0\ 10^{-9}$	$4.4\ 10^{-9}$

The collision risk value that result from the TOPAZ-TAXIR simulations is composed of risk contributions from combinatorially many event sequences (Stroeve et al., 2003). In particular, the event sequence classes include the status of technical systems, such as alerting systems and communication systems, aircraft types, and human operator situation awareness. Since the adaptations of TOPAZ-TAXIR in the integration process with Air-MIDAS all consider assumptions regarding the behaviour of pilots flying, it is interesting to compare the risk decomposition for a pilot flying in the original and modified models. In particular, in Figure 2, collision risk results are shown for the situations that

- the pilot flying of the taxiing aircraft believes to be on a regular taxiway, or
- The pilot flying of the taxiing aircraft believes that runway crossing is allowed.

In the first case the pilot is lost, in the second case the situation awareness corresponds well with the actual position of the aircraft. It can be observed in Figure 2 that in both the original and the modified model, the risk contribution for the situation that the pilot is aware to be on a regular taxiway exceeds the risk contribution for the situation that the pilot is aware to be on a runway crossing. However, the difference between those risk contributions is smaller in the modified model than in the original model.

On the one hand, the reduced difference in the risk contributions between the model versions is due to an increase in the risk contribution for the situation that the pilot flying is aware to be on a runway crossing in the modified model. The model modifications that may effect this risk increase concern the braking initiation times by the pilots flying of both aircraft, and the duration of the visual observation tasks of the pilots flying of both aircraft.

For a further evaluation of the effects of these modifications a sensitivity analysis is required. As a preliminary finding, the risk increase is especially due to the increase in mean braking initiation times and may be to a smaller extent due to the increase in mean visual observation time in the opportunistic mode.

On the other hand, the reduced difference in the risk contributions between the model versions is due to a decrease in the risk contribution for the situation that the pilot flying is aware to be on a regular taxiway in the modified model. The risk decrease in this situation is effected by all model modifications. The combined effect of the changes in the braking initiation times and the visual observation times is a risk increase. The decrease in the mean inter-monitoring time of the pilot flying of the taxiing aircraft leads to a risk decrease because it causes the pilot to monitors for conflicting traffic more often. The combined effect turns out to decrease risk.

Conclusions

The results showed that the Air-MIDAS based adaptation did lead up to a factor two reduction in assessed collision risk level. This result alone demonstrates that it is feasible and useful to couple Air-MIDAS and TOPAZ. More importantly this means that we have now running two human performance simulations for more or less the same situation. This gave us the unique chance to make further comparisons between the two simulation approaches.

We examined the change in collision risk assessment resultant from the integration of these two models. In the scenario examined, the actions of the flight crew and ATC are largely perceptual-motor response to runway incursion. The impact assessment reported reflects the change in those characteristics. More complex decision making or coordinated action among agents and safety augmentation technologies would require full representation of the models of those more complex interactions.

In order to recognize the logical pattern in these differences, one should be aware that both are aimed to assess quite different top-level metrics. Air-MIDAS top-level metric is the behavioral pattern of human operators; while TOPAZ top-level metric is collision risk. The implied focal attention in TOPAZ is on performance, error making and error propagation among multiple agents versus memory and task scheduling and performance in Air-MIDAS. For error mechanisms the error recovery model of Amalberti & Wioland (1997) has been reported for two types of stress levels. This is reflected by the two control modes of TOPAZ and avoids the need to model a lot of memory and task performance characteristics. In Air-MIDAS the adoption of the Skill Rule Knowledge (SRK) model of Rasmussen for task performance leads to three control modes, and with this the need to model memory and task scheduling and performance in detail. The complementarity of TOPAZ and Air-MIDAS makes it so interesting to compare simulation results obtained by both approaches.

From a validation perspective both approaches have much in common: they produce results on basis of carrying out simulations with a mathematical/computational model and by its very nature, a mathematical/computational model differs from reality. In order to validate a mathematical/computational model in a systematic way, the following activities should be performed:

- Identification of the differences between the mathematical model and the reality, and
- Assessment of the effect of these differences on the value of the output metric(s).

This validation process termed bias and uncertainty assessment is scheduled to be undertaken for the integrated simulations of Air-MIDAS and TOPAZ-TAXIR for the runway operation considered.

References

Amalberti R, Wioland L (1997), Human error in aviation, In: Aviation safety, pp. 91-108, H. Soekkha (Ed.), 1997.

Blom HAP, Bakker GJ Blanker PJG, Daams J, Everdij MHC, Klompstra MB (2001a). Accident risk assessment for advanced air traffic management, Eds: GL Donohue and AG Zellweger, Air transportation systems engineering, AIAA, pp. 463-480.

Blom HAP, Stroeve SH, Everdij MHC, Vander Park MNJ (2003), Human cognition perfromance model to evaluate safe spacing in air traffic, Human Factors in Aviation Safety, Ashgate, Vol. 3, pp59-82.

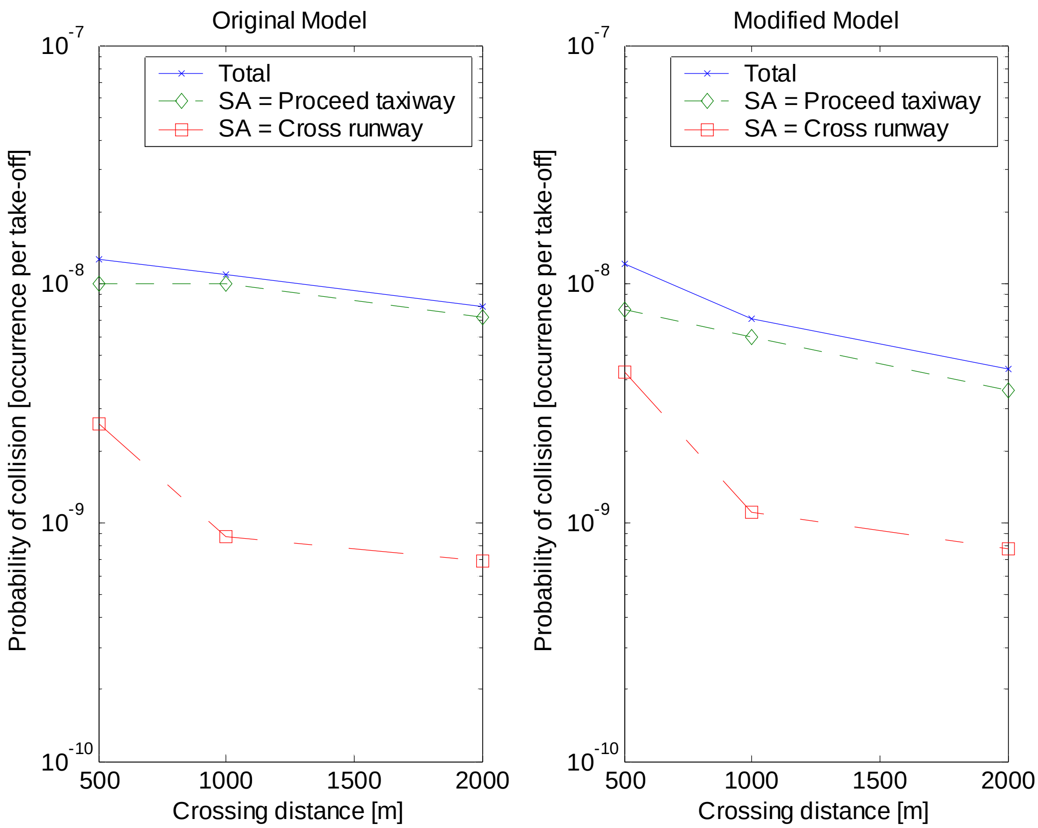


Figure 2: *Total risk per take-off aircraft to collide with a crossing aircraft and contributions to this by correct and incorrect SA by the PF of the crossing aircraft; left for the original, right for the modified TOPAZ-TAXIR model.*

Cacciabue PC (1998), Modelling and simulation of human behaviour in system control, Springer.

Corker, K. (2000), Cognitive Models & Control: Human & System Dynamics in Advanced Airspace Operations, Eds: N. Sarter and R. Amalberti, Cognitive Engineering in the Aviation Domain, Lawrence Earlbaum Associates, New Jersey.

Dekker, S. (2001). The Field Guide to Human Error Investigations. Cranfield University Press. Ashgate, London.

Everdij MHC, Blom HAP (2005), Piecewise Deterministic Markov Processes represented by Dynamically Coloured Petri Nets, Stochastics and Stochastic Reports, Plenum Press, forthcoming, February 2005.

Hollnagel E (1993). Human Reliability Analysis: Context and Control. Academic Press, London.

Kumamoto H, Henley EJ (1996), Probabilistic Risk Assessment and Management for Engineers and Scientists, IEEE Press.

Labeau PE, Smidts C., Swaminathan S. (2000), Dynamic reliability: towards an integrated platform for probabilistic risk assessment, Reliability Engineering and System Safety, Vol. 68, pp. 219-254.

Reason, J. (1990) Human Error. Cambridge University Press. Cambridge, England.

Stroeve SH, Blom HAP, Van der Park MNJ, (2003). Multi-agent Situation Awareness Error Evolution in Accident Risk Modeling, Proc. 5th USA/Europe Air Traffic Management R&D Seminar, Budapest, Hungary, June 2003.

Swain, A. D. & Guttman, H.E. (1983). Handbook of human reliability analysis with emphasis on nuclear power plant applications. Sandia National Laboratories, NUREG/CR-1278. Washington D.C.

Wickens, C.A. (1992), *Engineering Psychology and Human Performance*, second edition, Harper Collins, N.Y.

AB INITIO TRAINING IN THE GLASS COCKPIT ERA: NEW TECHNOLOGY MEETS NEW PILOLTS
A Preliminary Descriptive Analysis

Paul A. Craig, **John E. Bertrand**, **Wayne Dornan**, **Steve Gossett**, **Kimberly K. Thorsby**
Middle Tennessee State University
Murfreesboro, Tennessee

The Aerospace Department at Middle Tennessee State University and the NASA Langley Research Center entered into a cooperative agreement in 2003. The project is named the SATS Aerospace Flight Education Research (SAFER) and is part of NASA's Small Aircraft Transportation System (SATS) initiative. The SATS project envisions a future flight environment that employs small aircraft to transport people and cargo from point to point using smaller, under utilized airports instead of major gridlocked airports. The aircraft used in the SATS vision would take advantage of a range of emerging technologies including glass cockpits, new structures, and new engines. But with the understanding that the best aircraft and the best systems are still only as good as its operator, MTSU Aerospace set out to explore how pilot training might be different in the SATS environment. The SAFER project therefore takes beginner pilots and completes their initial Visual Flight (VFR) and Instrument Flight (IFR) flight training in technically advanced aircraft to determine how best to educate the next generation of pilots in the next generation of aircraft.

Introduction

Once the use of "glass cockpit" technology was reserved for airline and military flight crews. Today this technology can be purchased off-the-shelf from several general aviation aircraft manufacturers. Placing a general aviation pilot directly into such a sophisticated cockpit has many worried. The General Aviation Technically Advanced Aircraft (TAA)– Safety Study (2003) has already identified several accidents attributed to the fact that the pilots were not familiar with the technology available to them in their aircraft. Several studies are underway to aid pilots as they transition from round-dial airplanes to computerized flight displays – but that is not the emphasis of the study at MTSU. The SAFER project brings in potential pilots with little or no previous experience and teaches them to fly from the beginning with TAA.

The Students

All the students of the SAFER project are college students majoring in Aerospace at Middle Tennessee State University. To become eligible for the SAFER project students had to meet two criteria. First, they must have already been accepted into the program's flight laboratory, which requires a 2.5 cumulative college GPA, or a 2.8 high school GPA for incoming freshman students. Second, the students must have had less than five flight hours of experience with a flight instructor. Fifteen students formed the first cohort of SAFER students. The training began in September 2004 as the fall semester started. The second cohort began in January 2005 as the spring semester started.

The Training Syllabus

The features of the Garmin G-1000 system make it possible to blend the world of visual flight and the world of instrument flight – but that is not the traditional way that students are taught today. Students are taught visual flying first and pass a series of tests to obtain the Private Pilot Certificate. The Private Pilot then takes on additional training and testing to become Instrument Rated and this allows the pilot to fly in and through the clouds. The Primary Flight Display of the G-1000 provides a representation of the horizon that is far advanced from basic attitude gyro indications. The system, in effect, turns a dark night into daylight, and clouds into clear weather. The researchers wanted to take advantage of this capability and sought to teach the new students both the visual and instrument skills all at once.

Part of the cooperative agreement with NASA called for the SAFER project to work in conjunction with the FAA Industry Training Standards (FITS) initiative. The FITS group had previously developed a generic flight training syllabus that combined the training for both Private Pilot and the Instrument Rating into one. The SAFER team took the generic FITS combination syllabus and rewrote it for specific use at MTSU. In time, the syllabus was approved by the FAA under Part 141 and added to

MTSU's existing Air Agency Certificate. The MTSU version of the FITS syllabus (2004) became the first combination Private and Instrument Course for Technically Advanced Aircraft ever approved by the FAA.

The syllabus was unique in two other important ways. First, the entire combination Private and Instrument course is scenario based. Traditionally, pilots are trained using a series of maneuvers that the student masters with drill and practice. The SAFER syllabus still teaches basic skills, sometimes referred to as "stick and rudder" skills, but instead of drill and practice, the maneuver is incorporated into an overall scenario lesson. The very first lesson of the SAFER syllabus is a flight to another airport – a mission, rather than a set of maneuvers. The second unique feature of the SAFER syllabus is that it has no minimum flight time requirements. Traditionally trained students must meet several minimum flight time requirements to move from one step to another and to receive FAA pilot certification. It would be possible for a pilot to have achieved an acceptable performance level in a particular area of training, but still be required to take additional training just to reach the minimum flight time number. Students in the SAFER project are judged by performance only not flight time. When students complete each lesson of the SAFER syllabus they are recommended for testing regardless of how many or how few flight hours they have accrued.

The FAA Exemption

A major problem for the SAFER students is that they are training in a time of transition. The syllabus that they use and the airplane that they use are all new, but the FAA testing is old. Today, the Code of Federal Regulations 14, Part 61.65(a)(1) (2005) requires that an applicant for the Instrument Rating, already be the holder of the Private Pilot Certificate. But the SAFER syllabus bypasses the Private Pilot test when students would otherwise be eligible to take it. Instead, the SAFER students remain as student pilots until the day that they take the combination test and become Private Pilots and Instrument Pilots all at once. So the SAFER syllabus, is in fact, in violation of the Federal Aviation Regulations. To remedy this incongruency, the SAFER researchers petitioned the FAA for relief from 61.65(a)(1) and on December 10, 2004, the FAA granted an exception to this rule for the SAFER project. FAA exemption number 8456 (2004) allows the SAFER students to take a single practical test to gain both Private Pilot and Instrument Pilot privileges. The exemption came with a new Practical Test Standard (PTS) that is to be used by a pilot examiner when administering the combination test. The exemption has only been granted to MTSU and the SAFER project and extends until December 1, 2006.

The exemption has not eliminated all "old versus new" roadblocks to the training. The SAFER students still are required to take two knowledge tests that are administered via computer. The two tests contain questions that are not applicable to technically advanced aircraft. The new PTS that came along with the exemption is better than two separate tests, but still requires many drill-and-practice type maneuvers that do not match well with the SAFER scenario based syllabus. This forces the SAFER students to step out of the role of the scenario and occasionally revert back to pure maneuver practice simply to meet the requirements of the test. Using the old form of testing with the new form of training has become a very real impediment to the students that lengthens the time of training and pushes instructors to "teach to the test" rather than "teach for the real world" as the SAFER project intends to do.

The Methodology

The researchers of the SAFER project are in the preliminary stages of the data collection. The project is on going and the final report of findings will come at the conclusion of the project. The researcher are gathering data to help answer some of the basic research questions: If you teach people to fly from the very beginning using glass cockpits, are there any topics and/or skills that have been taught traditionally that are now no longer necessary? Will glass cockpits create new challenges for beginners that have not been contemplated previously? Can pilots learn essential skills faster and more completely using TAA? To help find some answers, the researchers started a comparison between the SAFER students and the performance of past students that were taught in traditional ways.

The Airplanes

In 2003, the Aerospace Department was able to purchase 25 new airplanes for their professional pilot degree program. Of these, eleven were

Diamond DA40s. As a part of the NASA cooperative agreement, five of the DA40s came to MTSU with the Garmin G-1000 glass cockpit system installed. These five airplanes were taken out of the traditional flight training fleet and are used exclusively within the SAFER project.

Early Findings

The researchers first looked backward to evaluate traditional flight training from the first flight until a person became an Instrument Rated Pilot. The pilot training records of past students served as archival data of traditional flight training. Nineteen past student training records were used in the study. Researchers took the training records of students who had taken both their Private Pilot and Instrument Pilot training all at MTSU and all used the traditional FAA approved syllabus. The traditional syllabus adopted by MTSU and approved by the FAA is the Jeppesen Private Pilot Syllabus (2002) and the instrument portion of the Jeppesen Instrument and Commercial Syllabus (2003). The two publications are commercially available and widely used as an industry standard throughout civilian flight training. The traditional path from first flights to Instrument Rated pilot goes first through the Private Pilot curriculum and testing, then through a series of visual flights to other airports (cross country), and finally to the specific training that leads to testing for the Instrument Rating.

Bottlenecks

Using the archival data provided by the FAA training records, the researchers examined the process of traditional training. What was discovered was a pattern of predictable bottlenecks throughout the training. A bottleneck, for this purpose, is defined as a lesson or area of training that requires the student to receive additional instruction, beyond that which is prescribed in the FAA syllabus, to reach mastery of that lesson or area. These bottlenecks represent areas that are more difficult for students, in that it requires more training to achieve the completion standards. One of the basic research questions is: Do the SAFER students experience the same bottlenecks in their training as traditional students do? Would SAFER students have less problems, or different problems than their counterparts who received the type of training that is available nationwide to the general public and to other college students? In order to answer this question the researchers first identified the traditional bottlenecks in the three phases of the training: Private Pilot, Cross Country, and Instrument.

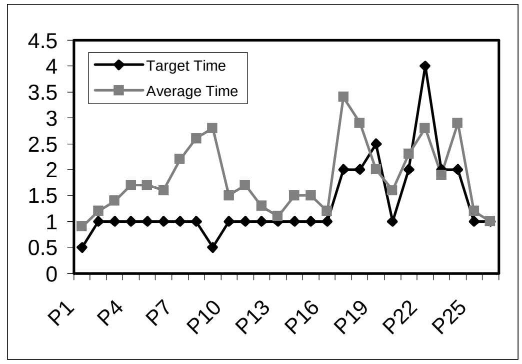


Figure 1. *Private Pilot Bottleneck. Flight Hours versus Lesson Numbers.*

Figure 1 illustrates the bottlenecks faced by traditional students during their Private Pilot training. The Target Time or recommended number of flight hours that should allow mastery in the topics and maneuvers contained in the lesson. The Target Time comes from the Jeppesen Private Pilot syllabus. The Average Time is the actual average hours it took for the traditional students to achieve mastery. It is clear that there are two predictable bottlenecks in this curriculum: Lessons 7 - 9, and Lessons 17 – 18. Lessons 7, 8, and 9 occur just prior to the students first solo flight. Lessons 17 and 18 cover cross-country navigation planning.

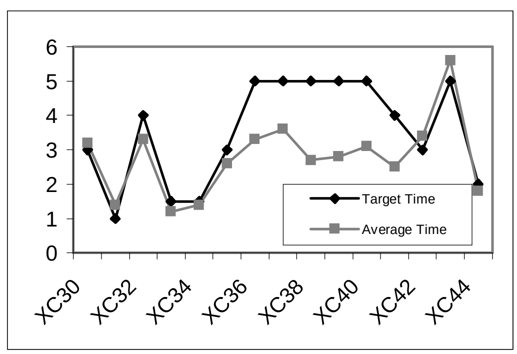


Figure 2. *Cross Country Bottlenecks. Flight Hours versus Lesson Numbers.*

Figure 2 illustrates the relationship between the target flight hours and the actual average time students needed in the cross-county phase. As Figure 2 indicates, students have few bottlenecks in this part of the curriculum. In fact, from Lessons 36 – 42, the students are actually flying less than prescribed. These lessons each require a flight to another airport with varying distances, but all greater than 50 nautical miles. One possible reason for the fact that average flight time is less than prescribed time in Lessons 39 through 42 is so students can make up for time overruns during the Private Pilot phase of training. If a student passes the Private Pilot tests with above average total flight time, this could be made up by undercutting the prescribed cross-country flight time.

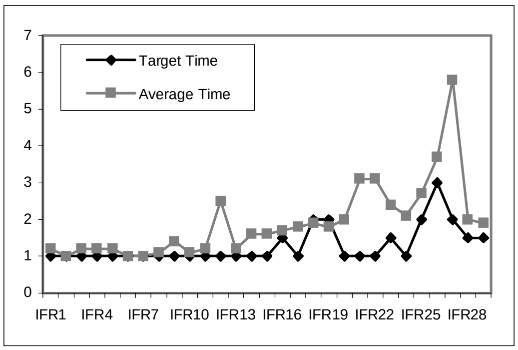


Figure 3. *Instrument Rating Bottleneck. Flight Hours versus Lesson Numbers.*

Figure 3 illustrates that last portion of the path to the Instrument Rating – the actual instrument training. Three bottlenecks are evident in the Jeppesen syllabus for instrument lessons: Lesson 12, Lessons 20 and 21, and Lesson 27. Lesson 12 contains the skill of VOR tracking and radial intercepting as well as partial panel tracking. Lessons 20 and 21 contain the ILS instrument approach, including the partial panel ILS. Lesson 27 is an instrument cross-country review flight.

Setbacks

Figures 1, 2, and 3 all illustrate the average number of flight hours that was required by students to reach mastery on that lesson. The researchers also observed the number of "setbacks" that a student experienced. A setback, in this case, is the need for a student to repeat a lesson that was previously flown. Among the archival data retrieved from the traditional student's training records, 449 setbacks were discovered. Of these, 77 setbacks took place just prior to the first solo flight – an area identified as a bottleneck in Figure 1. This number is 17.1% of all the setbacks experienced by traditional students. Setbacks continued for the traditional students throughout the remainder of the curriculum: 37.6% of the setbacks occurred during the Private Pilot and Cross Country phases of training past the first solo, and 45.2% of the setbacks took place within the instrument phase of the training. This tends to indicate that traditional students run into difficult lessons throughout the entire curriculum in all phases of Private, Cross Country and Instrument – there is never a time when it becomes "easier" for them.

First SAFER Student Data

Since the SAFER syllabus does not have minimum flight times for the course or for each lesson, there is no target flight time number to compare with actual flight time averages, as was the case with the traditional students' data. This makes a direct comparison between Traditional and SAFER student performance more difficult. Also, the Traditional students and the SAFER students do not come across the same topics in the same order, so a lesson-by-lesson comparison is also not direct. However, over the course of the SAFER syllabus, the same set of mastery skills are required, so an evaluation of student setbacks among the groups is possible.

The SAFER students within the first cohort experienced a total of 97 setbacks. Again, a setback is a repeated lesson. Lessons from both traditional and SAFER syllabi require a mastery of the subject matter before the student moves on to the next lesson, so a repeated lesson indicates that the student had difficulty with the subject matter contained in the lesson. Of the 97 setbacks, 59 took place among the SAFER students in the first nine, pre-solo lessons. This represents 60.8% of the total setbacks. The traditional students only had 17.1% of their setbacks occur during this portion of the curriculum.

	Traditional	SAFER
Pre Solo	77 of 449 17.1%	59 of 97 60.8%
Pvt & X-C	169 of 449 37.6%	15 of 97 15.4%
Instrument	203 of 449 45.2%	23 of 97 23.7%

Table 1. *Setback Percentages*

Table 1 presents the comparison of setbacks among the two pilot groups. The traditional students had far fewer setbacks in the early, pre-solo training, but their setbacks increase as they progress through the syllabus. The SAFER students had the greatest difficulty early on, but their setbacks diminished as they continued through the SAFER syllabus.

Skills Comparison

The lessons in the traditional curriculum produced student bottlenecks at Private Pilot lessons 7, 8, 9, and 17, and in the Instrument syllabus at lessons 12, 20, 21, 24, and 27. These lessons each contain many maneuvers and procedures embedded within each lesson, but there is a main area of lesson emphasis in each case. A bottleneck is an area in which students experience difficulty, so the main area of that lesson's emphasis would therefore be the source of that difficulty. Takeoff, landing, and emergency procedures present a significant challenge to all beginning flight students – especially landings. Evidence of this fact is shown by the bottleneck present with traditional students at lessons 7, 8, and 9, and by the disproportionately large number of setbacks at Lesson 9 for the SAFER students. This is the phase of flight where Traditional students out performed the SAFER students – see Table 1 where just prior to solo is where 60% of all SAFER setbacks took place and where only 17% of Traditional students setback took place. Beyond this phase of flight training however, the SAFER students reduced their number of setbacks precisely in areas where Traditional student hit bottlenecks.

On Lesson 17, Traditional students hit a bottleneck – see Figure 1. This area of emphasis is Cross Country Flight Planning. This lesson requires the student to obtain and assess weather information that is pertinent to a proposed visual flight. The student must plan a course of flight allowing for wind drift. The student must calculate time, speed, and fuel consumption for the flight and become extremely familiar with aeronautical charts that depict the terrain features that the flight will traverse. Many traditional students experience a setback at this point, requiring repeat lessons and often multiple repeated lessons. Among the Traditional students there was 0.75 setbacks per student on Lesson 17. In the SAFER syllabus, Lesson 11 is the first lesson in which Cross Country Flight Planning becomes the complete responsibility of the student. Note that SAFER students start conducting mission-oriented flights to other airports from Lesson 1, so at this point they have already been exposed to the elements of Cross Country Planning. SAFER students experienced very few setbacks – an average of only 0.18 setbacks per student on Lesson 11.

Holding patterns prove to be difficult for students when learning the basics of instrument flying. Figure 3 indicates a gap between the target flight time and the actual flight time required to master Holding Patterns at Lessons 14, 15, and 16. Traditional students had 1.06 setbacks per student through these lessons. SAFER students also had difficulty with Holding Patterns. SAFER Lessons 24 and 25 cover Holding Patterns and students on these two lessons had an average of 0.85 setbacks per student.

One of the two largest bottlenecks that faced the Traditional students in the Instrument phase of training took place at Lesson 20 – 22. Lessons 20, 21 and 22 require the student to meet completion standards in the skills of Instrument Landing System (ILS) approaches and Partial Panel Approaches. The ILS requires excellent finesse of the airplane and Partial Panel work requires excellent situational awareness. Eleven percent of all Traditional student setbacks occurred in these three lessons alone, producing an average of 3.2 setbacks per student. At Lesson 22 of the SAFER syllabus, students have been tracking the ILS localizer for several lessons, but Lesson 22 is where full ILS and Partial Panel approaches are among the completion standards. SAFER students had no setbacks on Lesson 22.

The final test of an instrument pilot's readiness is IFR Flight Planning. This requires the instrument pilot to plan and assess the weather, and the weather minimums. The pilot must calculate speed, time, and fuel consumption, but also plan on a flight to an alternate airport if the weather is unsuitable at the intended destination. The pilot must be able to file and later receive an IFR clearance and be able to expertly communicate with air traffic controllers all through the flight. Traditional students had a setback at this lesson with an average of 1.18 setbacks per student. The recommended amount of flight time to complete this lesson is 2.0 flight hours. Traditional students however took 5.8 hours, on average, to meet the completion standards of the lesson. In the SAFER syllabus, the IFR Flight Planning review lesson is number 26. No SAFER students had a setback on Lesson 26.

A comparison of average student setbacks across the entire curriculum reveals that SAFER students have more setbacks in the pre-solo phase than do the Traditional students. But Traditional students continue to have setbacks in rising numbers throughout, while SAFER students have a reduction in setbacks. Figure 4 illustrates the average number of setbacks among student for the Pre-solo lesson, the remainder of the Private and Cross Country training, and the Instrument Rating instruction.

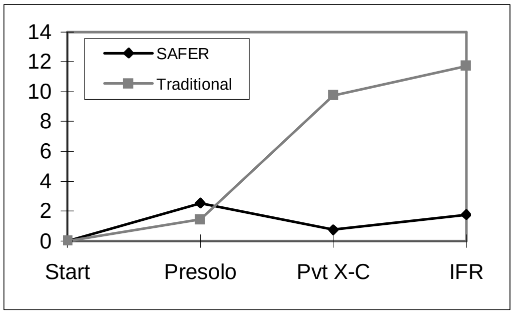


Figure 4. *Setbacks per student. Traditional students versus SAFER students.*

Conclusions

The researchers understand that we are dealing with small groups and that much more data must be taken before any claims can be made. But at this point the SAFER students have a greater number of setbacks in the lesson just prior to the first solo flight than do traditional students. The flight instructors that teach in the SAFER project say that the SAFER syllabus is very "front end loaded." This means that SAFER students are being taught cross-country flight planning, navigation, and instrument flight principles all before the first solo. The evidence, including Figure 4, seems to suggest that SAFER students pay a penalty for this expanded curriculum at the very start of the course. Traditional students are not taught cross country planning, navigation, and instrument principles before solo, and spend their time practicing takeoffs and landings in anticipation of the first solo. This focused attention on solo among traditional students may be why they perform with fewer setbacks in the pre-solo phase. But it appears that the "penalty" the SAFER students pay in the early lessons, are repaid later in the syllabus. The SAFER students seem to start reaping the rewards of their expanded curriculum after the first solo as the need for repeat lessons drops off to an average of only 0.76 setbacks per student between solo and the end of the SAFER stage 2 – which is approximately the cross country stage for Traditional student. Traditional students at this point experience an average of 9.73 setbacks. The evidence indicates that the largest benefit of the SAFER project is toward the end when both groups are preparing for the tests that cover the Instrument Rating. In that last phase of training the Traditional students had an average of 11.73 setbacks each, while the number of average setbacks among SAFER students was 1.76 each.

All the data presented here should be considered preliminary. The second SAFER cohort is underway at the time of this writing and the researchers will wait to see what additional data will bring to the conclusions. It is important to emphasize here that one of the overriding interest of the SATS program is to see if pilots can be trained in technically advanced aircraft that will meet or exceed the current training standards and to accomplish this in less time and with less money. The early information shows that the SAFER students who have completed the program and passed the combination Private Pilot and Instrument Rating test have done so with an average of 88.66 flight hours. The student who followed the traditional path completed the Instrument Rating at an average of 134.3 flight hours. The difference between the averages is approximately 45 hours. Forty-hours of flight instruction and airplane rental could cost the pilot approximately $6,000.

Although early in the project, the researchers are confident that the use of "glass cockpit" technology together with scenario training has great promise. Data from the remainder of the SAFER project will produce a list of "best practices" for flight instructors to use when teaching in TAAs. Ultimately, the project should lead to improvements and alterations to how pilots are to be trained in an environment of emerging technologies.

References

Code of Federal Regulations 14, Part 61.65(a)(1). (2005).

FITS Scenario Based Private/Instrument Pilot Certification Course Syllabus. (2004). Murfreesboro, TN: MTSU Aerospace.

Instrument Commercial Syllabus. (2003). Englewood, CO: Jeppesen Sanderson, Inc.

Private Pilot Syllabus. (2002). Englewood, CO: Jeppesen Sanderson, Inc.

Regulatory Docket Number FAA-2004-19151. Exemption Number 8456. (2004). Federal Aviation Administration.

General Aviation TAA Safety Study. (2003). Federal Aviation Administration.

COMPLEXITY MITIGATION THROUGH AIRSPACE STRUCTURE

M.L. Cummings, **C.G. Tsonis**
Massachusetts Institute of Technology
Cambridge, MA

D.C. Cunha
Instituto Tecnológico de Aeronáutica
São José dos Campos, SP, Brazil

Cognitive complexity is a term that appears frequently in air traffic control research literature, yet there has not been a significant distinction between different components of complexity, such as environmental, organizational, and display complexity, all which influence cognitive complexity. It is not well understood if and how these different sources of complexity add to controller cognitive complexity and workload. In order to address this need for complexity decomposition and deconstruction, an experiment was conducted to explore whether or not different components of complexity could be effectively measured and compared. The goal of the experiment was to quantify whether or not structure in airspace sector design, in combination with changes in the external airspace environment, added to or mitigated perceived complexity measured through performance. The results demonstrate that for a representative ATC task, the dynamic environment complexity source was a significant contributor to performance, causing lower performance scores. There was no apparent effect, either positive or negative, from increasing airspace structure represented through a display.

Introduction

Addressing the difference between environmental and innate human complexity (often referred to as cognitive complexity), Herb Simon describes an ant's path as it navigates across a beach. The ant eventually reaches its destination, but because the ant must constantly adapt its course as a result of obstacles, the path seems irregular, laborious, and inefficient. Simon points out that while the ant's path seems complex, the ant's behavior is relatively simple as compared to the complexity of the environment. Simon proposes the following hypothesis as a result, "Human beings, viewed as behaving systems, are quite simple. The apparent complexity of our behavior over time is largely a reflection of the complexity of the environment in which we find ourselves (Simon, 1981, p. 53)."

This distinction between innate or cognitive complexity and environmental complexity is especially relevant considering the considerable research conducted in air traffic controller cognitive complexity. Several studies have investigated air traffic control (ATC) information complexity issues (see Hilburn, 2004; Majumdar & Ochieng, 2002) for a review). In this literature, several common complexity factors have emerged to include traffic density, traffic mix, aircraft speeds, sector size, and transitioning aircraft. These factors are asserted to affect cognitive complexity. However, in light of Simon's ant parable, these factors really represent environmental complexity factors that influence cognitive complexity. This is an important distinction because as can be seen in Figure 1, there are several levels of complexity that can affect an individual's cognitive complexity level.

Figure 1 illustrates the decomposition of "complexity" as it applies to human supervisory control systems. Human supervisory control (HSC) occurs when a human operator intermittently interacts with an automated system, receiving feedback from and providing commands to a controlled process or task environment (Sheridan, 1992). In complex HSC systems, in general two layers of interventions, organizational and display design can exist to mitigate environmental complexity, and thus reduce cognitive complexity. Organizational interventions include goals, policies, and procedures such as separation standards, checklists, airspace structure, etc. For example, many airspace sectors are designed to promote predominant

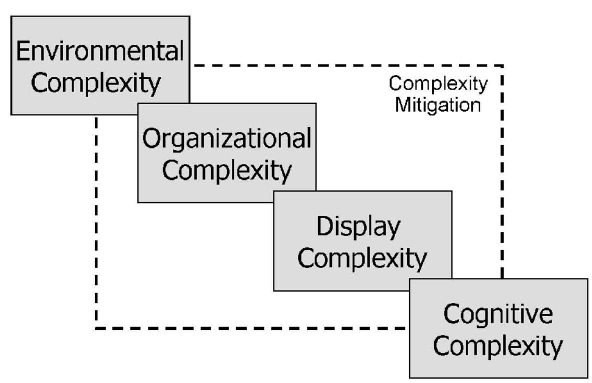


Figure 1: *Human Supervisory Control Complexity Chain*

traffic flows. Thus the design and the associated rules and procedures for control mitigate environmental complexity caused by increasing numbers of planes. However, when airspace becomes obstructed and saturated due to weather, congestion, etc., the need to follow procedures and sector limitations can over-constrain a problem, thus increasing the perceived complexity by the controller.

Displays are another example of intended complexity mitigation which could inadvertently add to complexity instead of reducing it. For air traffic controllers and in general all HSC operators, displays are critical in representing the environment so that a correct mental model can be formed and correct interactions can take place (Woods, 1991). In effect, to mitigate complexity, displays should reduce workload through transforming high-workload cognitive tasks such as mental computations into lower workload tasks through direct perception, i.e. visually (Miller, 2000). However, in complex and dynamic HSC domains such as ATC, it is not always clear whether a decision support interface actually alleviates or contributes to the problem of complexity.

Complexity and Structure

In addition to traffic density and related factors, it has also been hypothesized that the underlying airspace structure is a critical complexity factor (Histon et al., 2002). In theory, airspace structure provides the basis for mental abstractions which allows controllers to reduce complexity and maintain situation awareness. Histon et. al., (2002) propose that these mental abstractions, known as structured-based abstractions, can be generalized to standard flows (reminiscent of Pawlak's (1996) "streams"), groupings, and critical points. Providing air traffic controllers with these interventions, either explicitly through design or implicitly through policy, should help controllers improve through mental models, reduce overall complexity, as well as reduce perceived workload.

In a study investigating judgment and complexity, Kirwan et al., (2001) determined that airspace sector design was only second to traffic volume, in terms of contributing to cognitive complexity. In terms of the model in Figure 1, airspace sector design straddles both the organizational and display complexity categories. Designed by humans to mitigate environmental complexity, airspace structure is an organizational policy. However, airspace structure contains significant visual components represented in displays, thus it is an environmental complexity intervention both from an organizational and display perspective.

Including interventions in airspace sector design such as critical points (points through which aircraft must pass) and designated standard flows (such as jet ways) can increase order and improve predictability, and thus lower cognitive complexity. However, it is also possible that when uncertainty levels increase, usually as a function of dynamic environmental factors such as changes in weather and available airspace, these same airspace structures could actually add to complexity since a controller's mental model of the airspace design must be adapted to the new conditions. Airspace structure and procedures mitigate complexity in what are termed "nominal" situations, but when an "off-nominal" condition occurs, such as an emergency or unexpected weather phenomena, the resultant increasing uncertainty causes complexity to grow (Athenes, Averty, Puechmorel, Delahaye, & Collet, 2002).

While other research has attempted to quantify the individual elements of complexity as a function of traffic flow (Masalonis, Callaham, & Wanke, 2003), little attention has been directed towards understanding the different sources of complexity such as depicted in Figure 1. In addition it is not clear if and how these different sources of complexity add to controller cognitive complexity. In order to address this need for complexity decomposition and deconstruction, an experiment was conducted to explore whether or not elements of complexity as depicted in Figure 1 could be effectively measured and compared.

Method

Apparatus, Participants, and Procedure

To objectively investigate the impact of environmental and structural complexity factors on controller performance, a human-in-the-loop simulation test bed was programmed in MATLAB® (Figures 2 & 3). Since the subject pool consisted primarily of college students, it was necessary to devise a simplified and abstract task that addressed the aforementioned complexity concerns, but still represented fundamental elements of ATC. In a simplified en route task, subject controllers were assigned a single sector, and were only required to provide heading commands to aircraft, while velocities and altitudes were held constant.

Twenty egress areas were located in the periphery of the sector, and each incoming aircraft was assigned a specific egress point. The primary goal was to direct the aircraft (a/c) to the assigned egress, and when an aircraft exited correctly, a score was generated. To provide an incentive for flying through a pre-determined sequence of waypoints (representative of a flight plan), subjects could collect additional points by directing their a/c through these waypoints. The number of points that could be won at every waypoint was displayed. To discourage controllers from directing aircraft through unnecessary waypoints just to gain points, scores were penalized based on an aircraft's total time of presence in the airspace sector beyond that expected for the optimal pre-determined path. A final component of the overall score was the penalty for flying through a no-fly-zone. No-fly zones represented constrained ATC airspace such as thunderstorms, military operating areas, and prohibited areas. Example waypoints, optimal paths for particular ingress and egress points, and no-fly zones are represented in Figure 2. Maximization of total score was the subjects' goal, and their total score was displayed in real-time.

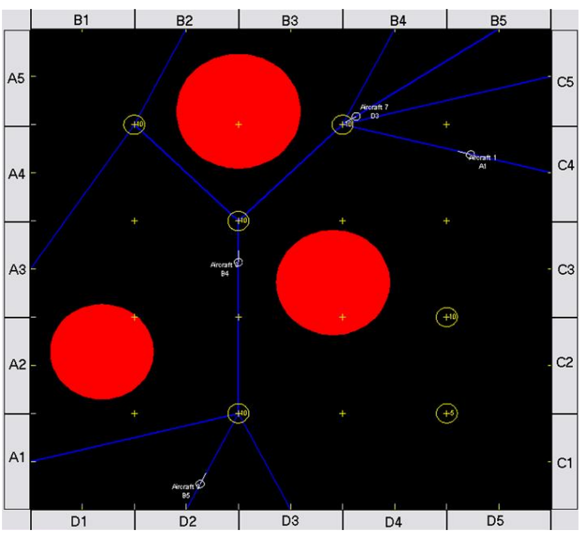


Figure 2: *Interface with optimal paths shown*

Training and testing were conducted using a Dell personal computer with a 21-inch color monitor, 16-bit high color resolution, and a 3.0GHz Pentium 4 processor. During testing, all user responses were recorded in separate files specific to each subject and scenario. A Visual Basic script was then written that scored and compiled the data into a single spreadsheet file for the subsequent statistical analysis. After signing required consent forms, subjects completed a tutorial that discussed the nature of the experiment, explained the context and use of the interface, and gave them the opportunity to understand the scoring mechanism. Subjects completed four practice scenarios that exposed them to every combination of independent variables. They then began the randomly ordered four test sessions, which also lasted until all aircraft had exited the airspace (approximately 6-7 minutes).

Experimental Design

Two independent variables were investigated. The first independent variable was the presence of structure, as displayed through the lines of maximum score (named "displayed structure"). As can be seen in Figure 2, in certain scenarios subjects were given structure through the display of the optimum paths (those that maximized the score as a function of waypoints and time). In the counter condition, subjects were given the waypoints (along with the number of available points), but were not shown the optimal path (Figure 3).

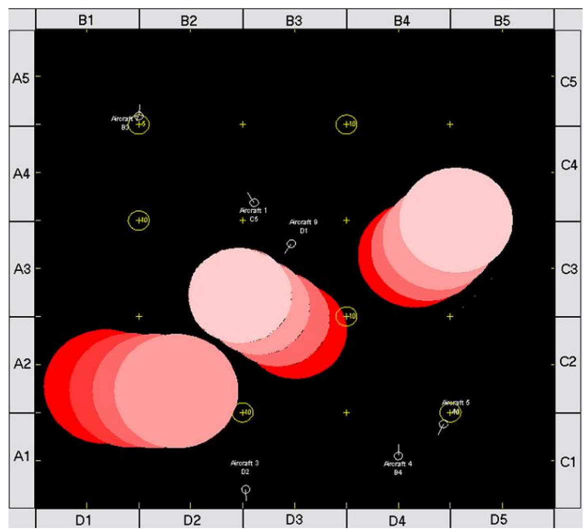


Figure 3: *Interface with dynamic no-fly-zones*

The second independent variable was the condition of the environment in terms of either static or dynamic no-fly-zones. In the dynamic condition, the no-fly-zones moved at rates of about two-fifths the aircraft velocity (figure 3), and representing changes in constrained airspace that often occur such as weather fronts and special-use airspace. It is important to note that the displayed lines were the optimum, but only in cases where they were not obstructed. In the dynamic condition, the dynamic no-fly zones cases would sometimes cover the paths, and thus the controller had to mentally regenerate new optimal paths. The motivation was to investigate whether or not such visual structure in an airspace sector, in combination with changes in the external airspace environment, added to or mitigated perceived complexity measured through performance.

A single dependent variable of total performance score was used. As described previously, the score was a linear and weighted function of aircraft egress correctness, bonus waypoints with penalties for no-fly-zone violations, and total time transitioning in sector. In the case of egress score, subjects received maximum points by directing their a/c to exit near the center of the egress, but did not receive points for exiting through the wrong egress. The egress scores decreased linearly from the center to the marked edges of the egress blocks. To maintain consistent scenario level of difficulty in order to minimize any learning effect, the four experimental scenarios were ninety degree rotations of each other. The statistical model used was a 2x2 fully crossed ANOVA and the four scenarios were randomly presented to a total of 20 subjects.

Results and Discussion

The 2x2 ANOVA linear model (with and without displayed structure and dynamic vs. static environment) revealed that for the performance dependent variable, the environment factor was significant ($F(1,74) = 54.55$), $p < .001$, all $\alpha < .05$). The displayed structure factor and the environment*displayed structure interaction were not significant. Figure 4 depicts the average performance scores across all four conditions. It can be seen on inspection that the performance scores were clearly higher in the static environment scenario as opposed to the dynamic environment phase. Whether subjects had less or more displayed airspace structure did not significantly affect their scores. These results demonstrate that for this representative ATC task, the environmental complexity factor was a significant contributor to performance, causing lower performance scores. There was no apparent effect, either positive or negative, from increasing airspace structure.

In terms of the model in Figure 1, this experiment demonstrated for this representative ATC task, the main component of complexity associated with controller workload was environment, and not organizational or display-related. Dynamically changing airspace structure was far more influential than the design of the airspace itself. Thus while sector design may be a contributing factor to air traffic controller performance, environmental complexity factors such as thunderstorms and special use airspace that intermittently becomes available, are significantly larger contributors to individual cognitive complexity.

These results provide quantitative support for previous subjective assessments of controllers that active special use airspace increases complexity and would benefit from some display intervention (Ahlstrom, Rubinstein, Siegel, Mogford, & Manning, 2001). In light of the results reported here, it is likely that special use airspace (SUA), an organizational constraint, can increase complexity for controllers not because of the actual structure of the airspace, because the status can change. When SUAs cycle between active and inactive, especially relatively rapidly, environmental complexity increases, and could negatively affect controller performance. Thus a by-product of an organizational policy could be increased complexity on the part of controllers.

These results indicate that the development of decision support tools to aid controllers in SUA management is an area of research that deserves more attention Because of the temporal and cyclic nature of SUA, possible design interventions could include some kind of timeline display for SUA scheduling as well as intelligent decision support agents that can predict in advance when airspace could become available or deactivated.

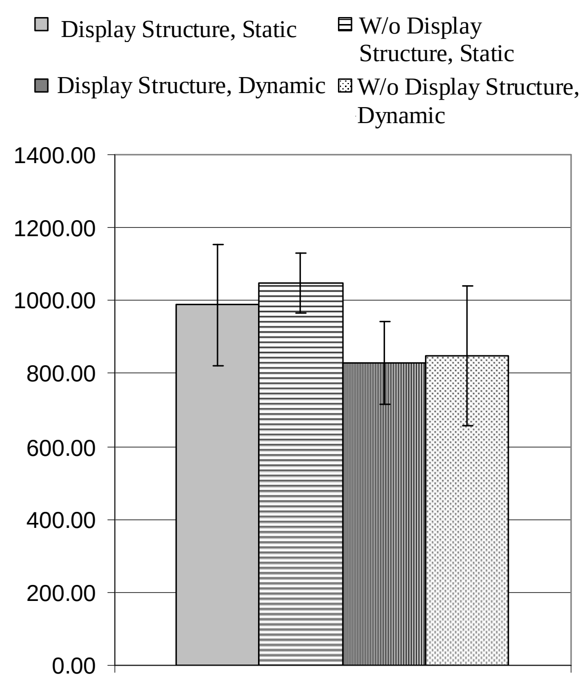


Figure 4: *Condition Mean Performance Times*

Conclusion

Complexity as it applies to the air traffic control environment cannot be simply categorized as "cognitive complexity," as there are different components of complexity, which are demonstrated in Figure 1. These components of environmental, organizational, and display complexity may not contribute in a linear and consistent manner to either cognitive complexity or performance. This study attempted to decompose two sources of complexity, an environmental factor caused by changing airspace, and an organizational/display factor caused by airspace design. Results show that the environmental complexity source of changing airspace was far more significant in influencing overall controller performance. These results support air traffic controllers' subjective opinions that special use airspace is a source of complexity (Ahlstrom et al., 2001), and that more work is needed for better display representation.

Acknowledgements

This research was sponsored by the Civil Aerospace Medical Institute.

References

Ahlstrom, U., Rubinstein, J., Siegel, S., Mogford, R. H., & Manning, C. (2001). *Display Concepts For En Route Air Traffic Control* (DOT/FAA/CT-TN01/06). Washington DC: U.S. Department of Transportation.

Athenes, S., Averty, P., Puechmorel, S., Delahaye, D., & Collet, C. (2002). *ATC Complexity and Controller Workload: Trying to Bridge the Gap.* Paper presented at the HCI-Aero.

Hilburn, B. (2004). *Cognitive Complexity in Air Traffic Control - A Literature Review* (EEC Note No. 04/04): Eurocontrol.

Histon, J. M., Hansman, R. J., Gottlieb, B., Kleinwaks, H., Yenson, S., Delahaye, D., & Puechmorel, S. (2002). *Structural Considerations and Cognitive Complexity in Air Traffic Control.* Paper presented at the 19th IEEE/AIAA Digital Avionics Systems Conference, Irvine, CA.

Kirwan, B., Scaife, R., & Kennedy, R. (2001). Investigating complexity factors in UK Air Traffic Management. *Human Factors and Aerospace Safety, 1*(2), 125-144.

Majumdar, A., & Ochieng, W. Y. (2002). *The factors affecting air traffic controller workload: a multivariate analysis based upon simulation modeling of controller workload.* Paper presented at the 81st Annual Meeting of the Transportation Research Board, Washington DC.

Masalonis, A. J., Callaham, M. B., & Wanke, C. R. (2003). *Dynamic Density and Complexity Metrics for Realtime Traffic Flow Management.* Paper presented at the 5th Eurocontrol / FAA ATM R&D Seminar, Budapest, Hungary.

Miller, C. (2000). The Human Factor in Complexity. In T. Samad & J. Weyrauch (Eds.), *Automation, Control, and Complexity: An Integrated Approach*. New York: John Wiley.

Pawlak, W. S., Brinton, C. R., Crouch, K., & Lancaster, K. M. (1996). *A framework for the evaluation of air traffic control complexity.* Paper presented at the AIAA Guidance Navigation and Control Conference., San Diego.

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. Cambridge, MA: MIT Press.

Simon, H. (1996). *The Sciences of the Artificial* (3rd ed.). Cambridge: MIT Press.

Woods, D. D. (1991). The cognitive engineering of problem representations. In J. L. Alty (Ed.), *Human-computer interaction and complex systems*. London: Taylor and Francis.

FLIGHT PROGRESS STRIPS IN TOWERS: FREQUENCY INDEX AND PERCEIVED PSYCHOLOGICAL BENEFITS

Andrew R. Dattel[1], **Brian R. Johnson**[2], **Francis T. Durso**[2], **Carla A. Hackworth**[1], **Carol A. Manning**[1]
[1]FAA Civil Aerospace Medical Institute, Oklahoma City, OK; [2] Texas Tech University, Lubbock TX

A team of subject matter experts (SMEs) observed tower air traffic controllers as they marked flight progress strips (FPSs) at ten facilities. SMEs tallied marks and actions controllers made at various positions during 332 twenty-minute observation periods. During many of the observation periods, one or more marks or actions were targeted for interviews. The benefits controllers perceived from making the targeted mark or action varied across position. The findings from this study will help engineers preserve the functional benefits received from paper FPS when designing electronic FPSs.

Introduction

A flight progress strip (FPS) is often a critical, if not required, tool that aids air traffic controllers in safely managing the movement of thousands of flights daily. Traditionally, an FPS is a paper strip demarcated by sections called blocks that include information about an aircraft such as call sign, type of aircraft, flight level, heading, route of flight, and destination. A controller uses the FPS to update information about a flight, communicate information to other team members, verify a procedure has been executed, and organize information.

A controller accomplishes these tasks by making marks on the FPS or altering the position of the FPS by offsetting or moving it in the strip bay. With the advent of electronic substitutions for paper across industries, it is important to determine the operational as well as the psychological benefits of using paper FPSs.

Substituting electronic FPSs for paper without examining how controllers use paper FPSs may preempt the benefits a controller experiences by using paper. For example, Luff, Heath, and Greatbatch (1992) claimed that paper is superior to electronic substitution for the following five reasons: ease of data entry, flexibility of data entry, option with data input sequences, document differentiation, and mobility. Additionally, Vortac, Edwards, Fuller, and Manning (1993) identified multiple potential cognitive benefits of using paper FPSs in an en route environment.

However, transferring to a paperless environment has benefits. For example, Vortac et al. (1993) showed that prohibiting the use of a writing utensil and preventing the movement of an FPS actually improved prospective memory. In addition, Vortac, Barile, Albright, Truitt, Manning, and Bain (1996) found that en route controllers actually preferred electronic FPSs.

The goal of this study was to explore the operational and psychological benefits of tower controllers' use of paper FPSs. Without this understanding, important considerations may be excluded from the design of electronic replacements (Vortac, Edwards, & Manning, 1994). Although previous studies examined en route controllers (Durso, Batsakes, Crutchfield, Braden, & Manning, 2004; Durso & Manning, 2003; Durso & Manning, 2002), tower controllers and en route controllers may differ in the usage and perceived benefits gained from paper FPSs. Therefore, an important consideration for designing electronic FPSs is to maintain not just the superficial benefits, but also preserve any functional benefits, if any, gained from paper.

Method

Observations and interviews were conducted at ten air traffic control towers across the United States. Data collection occurred at two facilities within a metropolitan area in five different regions of the country. The towers varied in volume of aircraft operations for each respective airport (small, medium, or large), number of runways, and configuration of runways (crossing, parallel, or angular).

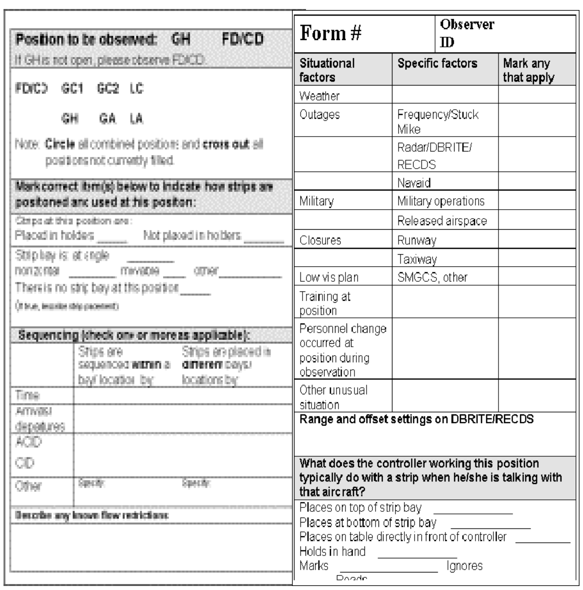
The four subject matter experts (SMEs) who conducted the observations were certified professional controllers (CPCs) who were not bargaining unit members. The SMEs had an average of 24.6 years experience as controllers. Researchers and graduate students of Texas Tech University and the Civil Aerospace Medical Institute (CAMI) conducted interviews. Two SMEs accompanied a group of interviewers to each metropolitan area, and no SME conducted observations at any facility that was in the same Federal Aviation Administration (FAA) region where that SME worked.

Participants

Observations were made of 332 positions across the ten facilities. Observations of 95 controllers who worked the flight data/clearance delivery (FD/CD), ground control (GC), or local control (LC) positions at the time of the observation were invited and agreed to participate in an interview. However, 175 interviews were conducted because several controllers were observed and interviewed on multiple occasions. The average age of controllers interviewed was 43.8 years, with an average of 17.8 years as a CPC, and approximately 10 years working at their respective facility.

Materials

During an orientation and training session prior to the first data collection, the researchers and SMEs developed an observation form. The first part of the observation form recorded situational factors such as current conditions, amount of traffic, and positions active at the time of the observation session (See Figure 1a). The second part of the observation form recorded the most likely actions and events a controller would mark on an FPS (See Figure 1b). The events marked were organized to represent the most logical flow.


Position to be observed: GH FD/CD
If GH is not open, please observe FD/CD.

FD/CD GC1 GC2 LC
GH GA LA

Note: **Circle** all combined positions and **cross out** all positions not currently filled.

Mark correct item(s) below to indicate how strips are positioned and used at this position:

Strips at this position are:
Placed in holders _____ Not placed in holders _______
Strip bay is: at angle ________ horizontal ________ movable ____ other__________
There is no strip bay at this position_____
(If true, describe strip placement)

Sequencing (check one or more as applicable):

	Strips are sequenced **within** a bay/ location by:	Strips are placed in **different** bays/ locations by:
Time		
Arrivals/ departures		
ACID		
CID		
Other	Specify:	Specify:

Describe any known flow restrictions:

Form #	**Observer ID**	
Situational factors	**Specific factors**	**Mark any that apply**
Weather		
Outages	Frequency/Stuck Mike	
	Radar/DBRITE/ RECDS	
	Navaid	
Military	Military operations	
	Released airspace	
Closures	Runway	
	Taxiway	
Low vis plan	SMGCS, other	
Training at position		
Personnel change occurred at position during observation		
Other unusual situation		

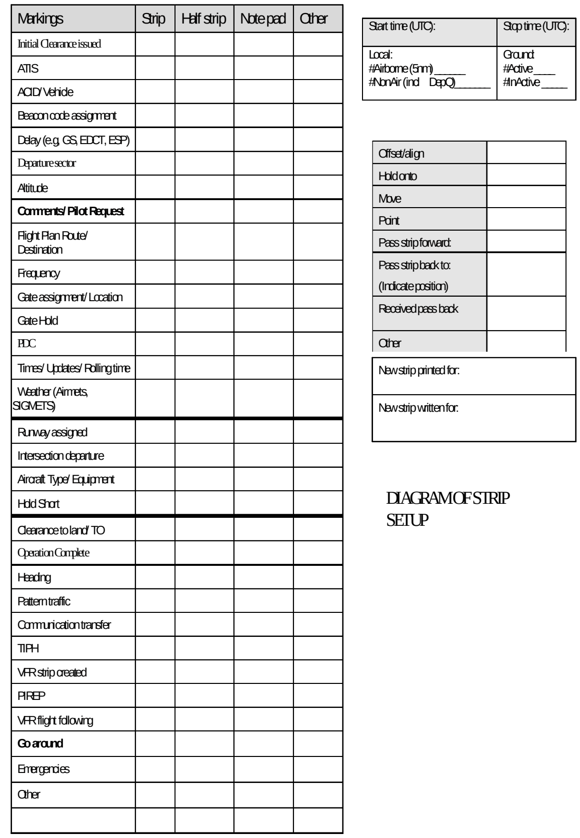
Range and offset settings on DBRITE/RECDS

What does the controller working this position typically do with a strip when he/she is talking with that aircraft?

Places on top of strip bay ___________
Places at bottom of strip bay ___________
Places on table directly in front of controller ________
Holds in hand ___________
Marks _______________ Ignores
Reads

Figure 1a. *Front side of observation form.*

Columns on the observation form indicated if the controller was using a strip, half strip, or notepad to make the mark. A half strip is a regular strip that has been cut in half when additional information is not needed for that flight. For example, many half strips are used for visual flight rule (VFR) traffic. The observation form also provided a space for SMEs to indicate how the controller handled the strip, including how it was placed in the strip bay and when it was passed to another controller. A notepad is a pad of paper the CPC uses to make notes or marks for flights that do not have an FPS.


Markings	Strip	Half strip	Note pad	Other
Initial Clearance issued				
ATIS				
ACID/ Vehicle				
Beacon code assignment				
Delay (e.g. GS, EDCT, ESP)				
Departure sector				
Altitude				
Comments/ Pilot Request				
Flight Plan Route/ Destination				
Frequency				
Gate assignment/ Location				
Gate Hold				
PDC				
Times/ Updates/ Rolling time				
Weather (Airmets, SIGMETS)				
Runway assigned				
Intersection departure				
Aircraft Type/ Equipment				
Hold Short				
Clearance to land/ TO				
Operation Complete				
Heading				
Pattern traffic				
Communication transfer				
TIPH				
VFR strip created				
PIREP				
VFR flight following				
Go around				
Emergencies				
Other				

Start time (UTC):	Stop time (UTC):
Local: #Airborne (5nm)_____ #NonAir (incl DepQ)______	Ground: #Active____ #InActive_____

Offset/align	
Hold onto	
Move	
Point	
Pass strip forward:	
Pass strip back to: (Indicate position)	
Received pass back	
Other	

New strip printed for:

New strip written for:

DIAGRAM OF STRIP SETUP

Figure 1b. *Back side of observation form.*

In addition to some basic biographical information, the interview form solicited open-ended questions a) if a targeted mark or action was required by the facility's standard operation procedures (SOP), b) the benefits received from making that mark or action, and c) if and how that mark or action helped achieve a goal. The interview form also consisted of 12 questions with a 7-point Likert scale about how much the targeted mark or action related to five psychological dimensions: communication, memory, workload, situation awareness, and organization.

Procedure

Data were collected over three days at two facilities in a metropolitan area. During two of the days, the SMEs rotated between each facility and for one day, the SMEs paired up at the larger of the two facilities. Thus, data were collected during all three days at the larger facility and during two days at the smaller facility. Therefore, each SME spent two days collecting data at the larger facility, and one day collecting data at the smaller facility.

Coordination was arranged through FAA headquarters, the regional offices, and individual facility managers to conduct the observations. Controllers were informed of the intent to collect data and had the opportunity to refuse being observed. Some controllers who agreed to be observed were asked to participate in an interview. The facility management agreed to allow the controllers to be interviewed without encroaching on their normal break time.

All positions, including positions that were always open and those that were periodically open, were observed using countered-balanced schedules. Because some positions may not have been open during the observation period, a backup position that was always active, such as FD/CD, GC, or LC, was randomly selected. If the controller at the selected position did not want to be observed, the SME selected the next position in the queue to be observed. This current paper analyzes data collected from the positions of FD/CD, GC, and LC when they were not combined with any other position.

During each observation period, the SME observed the controller at the selected position in an unobtrusive manner. SMEs were not "plugged-in." That is, the SMEs did not listen to the dialogue between controllers and pilots. For each observation, the SME tallied the controller's marks and actions on the standardized observation form. The SME also noted where the controller placed the mark (e.g., strip, half-strip, note pad) and what specific actions the controllers made (e.g., passing the strip to another controller, repositioning the strip). Each SME made twelve 20-minute observations each day. Generally, the SMEs conducted six consecutive observations before taking a one to two-hour break.

For most observations, the SME invited the controller for an interview because he or she made one or more marks or actions during the observation period. The SME handed the controller a receipt and encouraged him or her to talk with one of the interviewers at the controller's earliest convenience. Because the goal of this study was to provide a broad spectrum of how controllers use and benefit from marks and actions, SMEs were given the latitude to decide which marks and/or actions to select for an interview. On some occasions, the SME selected a typical mark or action; on other occasions, the SME selected an unusual mark or action. After an observation session, the SME gave a duplicate copy of the receipt, the observation form, and a copy of the strip or notepad containing the targeted mark or action to the interviewer.

After a controller selected for an interview was relieved from the controller's position, he or she met with an interviewer for approximately 15 minutes. The controller's receipt was matched with the interviewer's receipt to confirm that the controller had made the targeted mark or action that prompted the interview. In the event the controller stated he or she did not make that mark or did not remember the situation, another mark on the form was selected for interview. If several controllers had already been interviewed about the targeted mark or action, the interviewer selected another mark or action. The interviewer and controller then reviewed the targeted mark or action about when and why it was made. The interviewer asked the controller to consider that targeted mark or action when completing the questions on the interview form.

Results

Observation Form

An overall frequency count of marks made from FD/CD, GC, and LC for all facilities showed that "operation complete" was the most frequent type of mark or action made. Figure 2 shows the percentage of marks made by type of mark.

These percentages were broken down by position: FD/CD, GC, and LC. The frequency was the average number of marks per observation period (e.g., 20 minutes). Inspection of Figures 3 (a-c) shows how the frequencies of marks varied across positions. For example, "initial clearance" was rated in the top two types of marks made by FD/CD but at the bottom for LC. Likewise, "clearance to land" was rated as the third most frequent mark made by LC but the least frequent mark made by FD/CD and GC. Inspection of Figures 3(a-c) show FD/CD made more marks than either GC or LC; as expected, more marks were made at the larger facilities than at medium and smaller facilities.

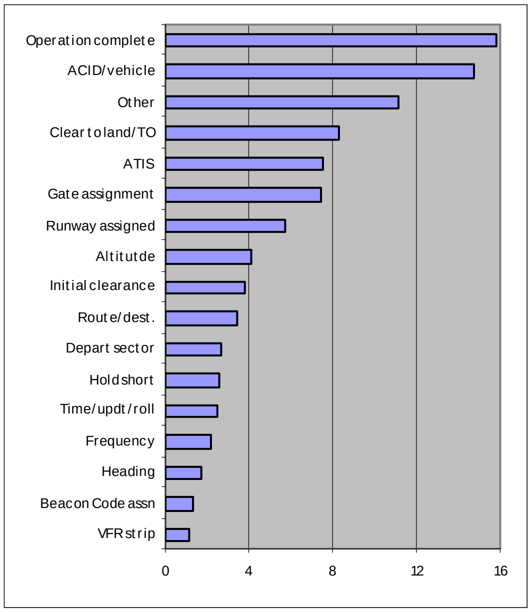


Figure 2. *Percentage of each type of mark made*

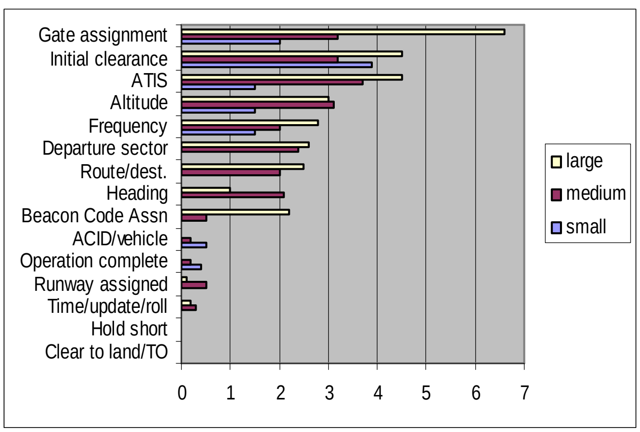


Figure 3a. *Average number of marks per 20-minute period for FD/CD.*

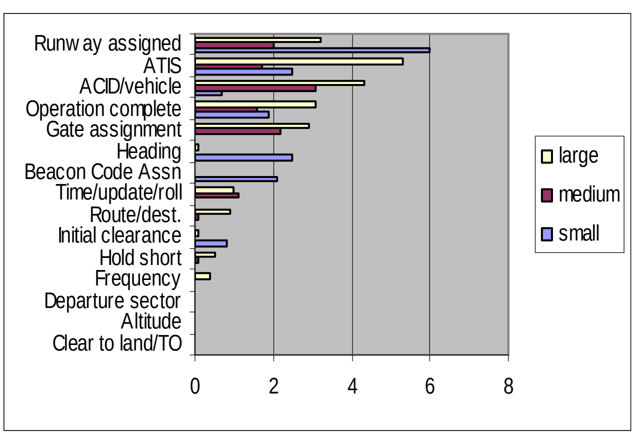


Figure 3b. *Average number of marks per 20-minute period for GC.*

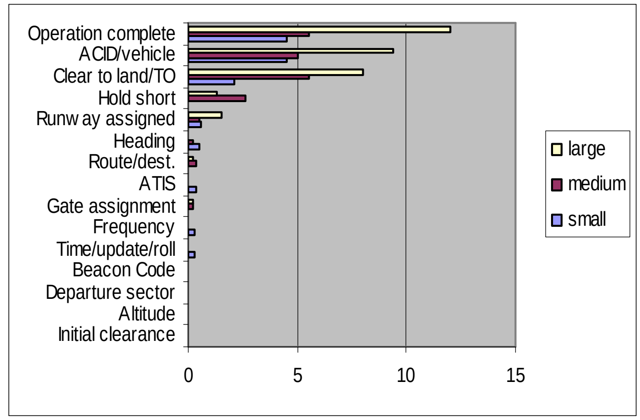


Figure 3c. *Average number of marks per 20-minute period for LC.*

Interview Questions

The first interview question asked the controllers if the targeted mark was required, benefited him or her, or both. Figures 4 (a-c) show the reason marks were made across positions.

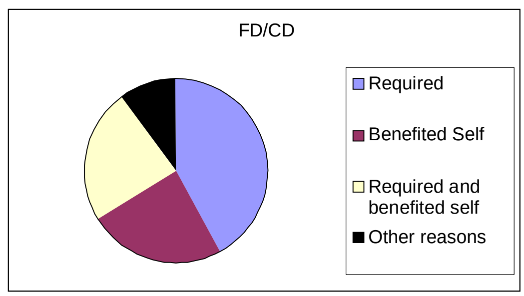


Figure 4a. *Reasons controllers at the FD/CD positions said they made targeted mark or action.*

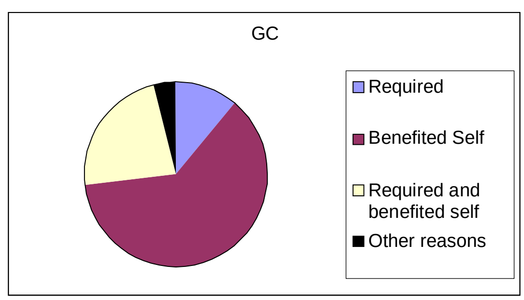


Figure 4b. *Reasons controllers at the GC position said they made targeted mark or action.*

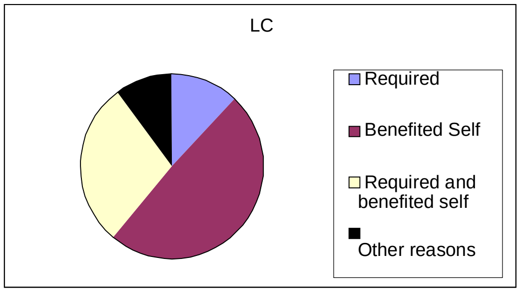


Figure 4c. *Reasons controllers at the LC position said they made targeted mark or action.*

Controllers at the FD/CD position were much more likely to make the targeted mark or action because it was required, rather than because of a perceived benefit to him or her. However, controllers at the GC and LC position were two times more likely to state that they made the mark because it benefited them, rather than because it was required.

The goal the controllers felt the targeted mark or action helped them achieve was scored and categorized as either psychological, operational, or both. An answer that was scored as psychological would be some individual benefit to the controller such as prospective memory or situation awareness. An answer that was scored as operational would be to satisfy a task required in the SOP, such as a count of aircraft or to communicate information to another team member. Figure 5 shows that the benefits for FD/CD were more operational than psychological and about evenly split for GC and LC.

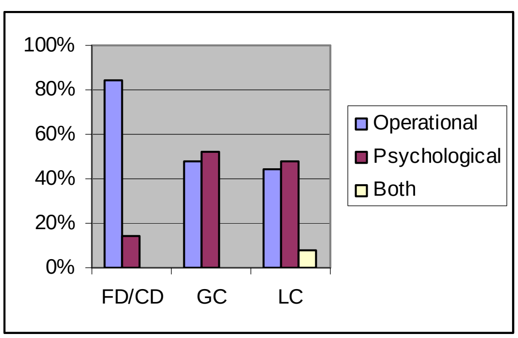


Figure 5. *Goals the marks or actions helped controllers achieve across positions.*

During the interview, controllers who said a targeted mark or action benefited them (rather than simply being required) were asked to specify the benefit. Their answers were scored along five psychological dimensions: communication, memory, organization, situation awareness, and workload. Figure 6 shows that the marks or actions made by FD/CD were most beneficial as an aid to communications and memory. GC made the marks as an aid to memory, situation awareness, and workload; LC made the marks and actions as an aid to memory and situation awareness.

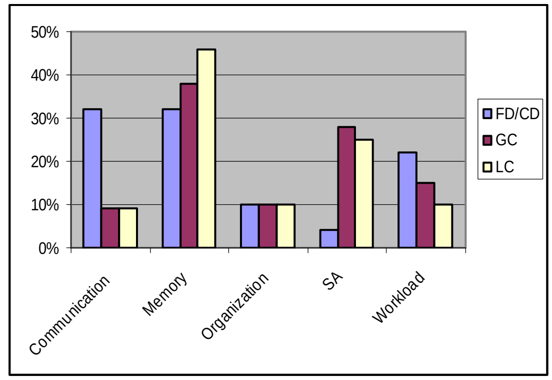


Figure 6. *Psychological benefits received form making targeted marks and action across position.*

For the final part of the interview, controllers used a 7-point Likert scale to answer questions about how well they felt the targeted mark or action related to each question. Each question was designed to elicit one of the five psychological dimensions. Table 1 shows mean ratings for controllers at all positions and separately by each position described above.

Psychological Dimension	Total Mean (SD)	FD/CD Mean (SD)	GC Mean (SD)	LC Mean (SD)
Communication	4.81 (2.27)	5.95 (1.67)	4.99 (2.13)	3.29 (2.16)
Memory	5.23 (1.90)	4.66 (1.91)	5.43 (2.07)	5.80 (1.62)
Organization	5.06 (1.75)	4.44 (1.81)	5.41 (1.85)	5.52 (1.59)
Situation Awareness	4.80 (1.86)	4.03 (1.90)	5.21 (1.85)	5.35 (1.59)
Workload	5.20 (1.74)	5.93 (1.44)	5.58 (1.54)	4.37 (1.91)

Table 1. *Ratings on 7-point Likert scale of psychological dimensions across position.*

Discussion

Some of the findings from these data are not surprising. For example, controllers at larger facilities made more marks than at smaller facilities, and certain marks were more specific to position. However, the benefits the controller gained from the marks varied by position.

Although controllers working the FD/CD position made more marks than did controllers working GC or LC, their marks overwhelmingly were made for operational reasons. The primary benefits of marks made by FD/CD were for communication. The FD/CD position updates the flight plan printed on the FPS, coordinates flight plan information with pilots, and starts the flow among the other team members. FD/CD does not direct the activities of surface or air movement. Rather, the primary responsibility of FD/CD is to ensure the flight plan is accurate and make sure this information is communicated between the flight crew and other CPC positions. Thus, other than communication and some workload benefits, strip marking seems to have few other benefits for the FD/CD position.

GC and LC experienced greater psychological benefits than FD/CD. The GC position is responsible for movement of aircraft and vehicles on the surface (e.g., taxiways). The LC position controls aircraft on active runways and during take offs and landings. The benefits most often perceived by GC were memory, workload, and situation awareness. The benefits most often perceived by LC were memory and situation awareness.

Additional analyses are being conducted using these data. Controllers' perceived benefits are being compared with importance ratings. The SMEs and a group of controllers not participating in this study independently rated the marks based on their perceived importance.

Before designing an electronic FPS, the perceived benefits found in this study need to be explored as actual benefits to the controller. Future studies will examine information requirements associated with paper FPSs and how they might be incorporated into electronic FPSs.

Acknowledgements

The authors would like to thank our sponsor, Steve Cooley (ATO-T), Dino Piccione, Human Factors Program Manager (ATO-P), and our SMEs, Rick Dillbeck, Ron Hubbard, Bob Hutson, and Ric Wunn.

References

Durso, F.T., Batsakes, P. J., Crutchfield, J. M, Braden, J. B., & Manning, C. A. (2004). The use of flight progress strips while working live traffic: Frequencies, importance, and perceived benefits. *Human Factors, 46*, 32-49.

Durso, F. T., & Manning, C. A. (2003). Following the paper trail: Design clues from paper flight strips, *Air Traffic Management Conference*, Budapest, Hungary.

Durso, F. T., & Manning, C. A. (2002). Spinning paper into glass: Transforming flight progress strips. *Human Factors and Aerospace Safety, 2*, 1-31.

Luff, P., Heath, C., & Greatbatch, D. (1992). Tasks-in-Interaction: Paper and screen based documentation in collaborative activity. *Conference Proceedings on Computer Supported Cooperative Work*, 163-170.

Vortac, O. U., Barile, A. L., Albright, C., Truitt, T., Manning, C. A., & Bain, D. (1996). Automation of flight data in air traffic control. In D. Herrman, C. McEvoy, C. Hertzog, P. Hertel, & M.K. Johnson (Eds.), *Basic and Applied Memory Research, Vol. 2.*, Mahwah, NJ: Erlbaum, 353-366.

Vortac, O. U., Edwards, M. B., Fuller, D. K. & Manning, C. A. (1993). Automation and cognition in air traffic control: An empirical investigation. *Applied Cognitive Psychology, 7,* 631-651.

Vortac, O.U., Edwards. M. B., & Manning, C. A. (1994). Analysis of complex sequential data: A case study involving air traffic control. *Human-Computer Interaction, 9*, 319-343.

PILOT SUPPORT FOR DISTANCE-BASED IN-TRAIL FOLLOWING TASKS

J.M.C. (Sjoerd) de Groot, **M.M. (René) van Paassen**, **Max Mulder**
Delft University of Technology, Faculty of Aerospace Engineering,
Control and Simulation Division
Delft, the Netherlands

Transferring the spacing task from the air traffic controller to the pilot can benefit efficiency and capacity. To separate a chain of aircraft, time-based rather than distance-based principles are preferred as they result in better performance in case of gradual reducing speeds in arrival streams. The present-day air traffic management systems, however, operate mainly on a spatial rather than a temporal basis, and air traffic controllers monitor the distance between trailing aircraft to determine if separation requirements are satisfied. If the disadvantages of distance-based spacing can be dealt with, the implications of introducing distance-based procedures for the current controller and pilot working environment would be much smaller than compared to time-based procedures. This paper presents the *spacing reduction concept* as a solution for the principal disadvantage of distance-based in-trail following, the slow-down effect. Displays and procedures were tested in a pilot-in-the-loop experiment. It is shown that distance-based spacing procedures can produce a stable chain of up to five aircraft, with very low pilot workload.

Introduction

Sequencing aircraft on an arrival route requires the air traffic controller to provide each aircraft steering commands, including speed, altitude and heading directions. Controllers attempt to have aircraft follow similar speed profiles along the arrival. When limits for separation are (to be) violated, speed clearances are issued to counteract the violation. In doing so, the controller transforms the 'global' mental picture of the approach sequence into a set of 'local' commands for one particular aircraft, a task that results in considerable workload. Transferring the spacing task from the controller to the pilot, i.e., *in-trail self-spacing*, would relieve controllers from this task, to the potential benefit of efficiency, capacity, and safety (Hoffman et al. 1999, Abeloos et al, 2001).

Pioneering work showed that spacing can be either *time-based* or *distance-based* (Sorensen & Goka, 1983, Williams, 1983). Generally, time-based spacing is preferable over distance-based spacing. A 'constant distance' criterion results in a slow-down in the speed-profile of a chain of aircraft because it requires trailing aircraft to fly the same ground speed as the leading aircraft. This is referred to as the *slow-down effect*. Time-based procedures would require the time distance between aircraft to be kept constant throughout the arrival. Subsequent lower speeds would not result in slow-down effects because the time requirement imposes subsequent lower spacings. These procedures, however, differ considerably from the way controllers and pilots currently operate. Current day radar systems and procedures operate under *spatial* representations of the air-traffic situation. Pilots share these problems during time-based self-spacing procedures, as the main sources of traffic-related information in the cockpit, like the Cockpit Display of Traffic Information (CDTI), also provide spatial situation presentations. Time-based procedures require new displays and tools to help pilots and controllers in handling time-based procedures (Lee et al, 2003).

Distance-based self-spacing procedures require far less modifications of procedures and systems. This paper describes how these procedures can be defined without the slow-down effect to occur. The pilot interface was developed simultaneously with the procedure (Pritchett & Yankoski, 2003). The results of an experiment are presented.

Spacing Reduction Procedure

With the 'constant distance' method the pilot task is to maintain a certain distance behind another aircraft, the *spacing requirement*. The aircraft that is being followed is called the target aircraft or *target*, the aircraft following target is called the own aircraft, or *own*. The spacing between aircraft is defined along the track. The difference between the required spacing and the actual spacing is the *spacing error*.

A requirement of self-spacing procedures is that every aircraft in a chain must follow the same ground speed profile. 'Basic' distance-based spacing does not automatically yield the same speed profile of the target aircraft like time-based methods do. It can be adapted, however, to bring about the same behavior. The crux of the matter lies in the fact that a *fixed* spacing requirement forces trailing aircraft to fly the same speed as the very first aircraft in the chain. But when the spacing requirement is allowed to *vary*, in a structural and procedurally well-described manner, along the approach, the slow down effect can be eliminated. This is shown in Figure 1.

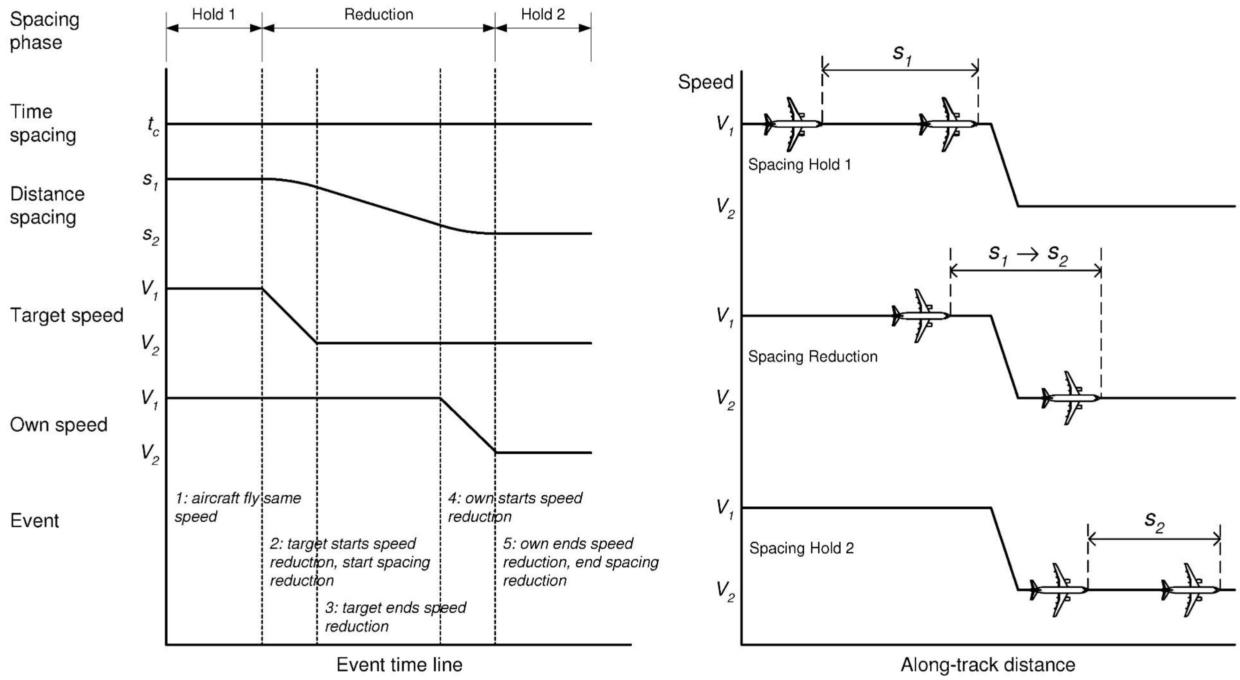


Figure 1. *Spacing development caused by a speed reduction of target from V_1 to V_2, with a constant time spacing t_c.*

The very moment at which the spacing requirement should change, is when *target starts its reduction in speed* (event 2). The change in spacing requirement should be discrete rather than continuous, and should be s_2. This s_2 belongs to the speed that target is heading at, V_2 in Figure 1. Right after the spacing requirement has been changed, *no action by own* is initially necessary. Only after the spacing has reduced from s_1 to s_2, own is to reduce speed to V_2. During the spacing reduction own is to *remain its current speed* V_1. The speed reduction to V_2 (event 4) is to take place some time before the spacing has reduced completely (event 5), because during the deceleration from V_1 to V_2 the spacing will still reduce a little. One can see that although the *speed profile* of own and target do not match exactly, the spacing requirement is met. The phase in which a speed reduction of the target aircraft initiates a spacing reduction, until the own aircraft meets the spacing requirement, is referred to as the *spacing reduction* phase. The phase during which the spacing is constant is referred to as the *spacing hold* phase. These phases can also be identified in Figure 1.

Chains

The present study will consider a chain of aircraft flying using the spacing reduction procedure. A leading aircraft receives speed clearances from the controller, the trailing aircraft operate under self-spacing procedures. Each trailing aircraft executes the self-spacing task with respect to its predecessor in the chain. All aircraft in the chain fly the same trajectory. The speed profile that is flown by the first aircraft in the chain is the *nominal speed profile* of a chain. Since self-spacing is best performed when aircraft follow the same speed profile, good self-spacing behavior should result in trailing aircraft flying speeds close to the nominal speed profile.

Controller Tasks

The task for the controller is to issue speed clearances to the very first aircraft in the chain. By doing this the controller defines the nominal speed profile for the chain that this aircraft is leading. A trailing aircraft should now be issued spacing clearances at the very moments that the target of this aircraft reduces speed.

Pilot Tasks

Two tasks exist for the pilot, dependent on the spacing phase. First, in the case of spacing hold, the pilot is to maintain a constant distance behind the target aircraft. Basically the pilot needs to adjust his speed so that the spacing does not change. Therefore own's *ground speed* has to be the same as target's ground speed. Spacing errors should be counteracted by changes in speed. The second task is executed when the pilot enters the spacing reduction phase. When the spacing requirement changes, the procedure requires the pilot to maintain the current speed V_1 until speed has to be reduced to meet the spacing requirement. After a change in spacing requirement, the spacing error and the closure rate both instantly increase. The pilot is expected, however, to *take no action* to counteract these 'errors'. Instead the pilot should be aware that the spacing error will decrease *by itself*, since the target aircraft has reduced speed. Only when the spacing requirement is met the pilot is to take action by reducing speed. An experiment will evaluate whether this procedure is acceptable.

Display Design

It is assumed that aircraft are equipped with ADS-B. The ADS-B message contains state information of the target aircraft, such as indicated airspeed, ground speed, track and position. This information together with the state of own makes it possible to calculate for example relative speed and distance information. The navigation display (ND) that stands at the basis of the experiment is a Boeing 747-400 ND, Figure 2. A design objective was to keep additional self-spacing systems as straightforward as possible. No automation and only simple algorithms are used.

Self-Spacing Symbols

Self-spacing augmentations included the target state information (speed, altitude), relative information (current distance), trend information (closure rate), intent information (target V_{cmd}) and predictive information (spacing capture marker, speed reduction counter). Also the spacing requirement with the allowed error margin was depicted on the display (spacing marker).

Traffic Symbols

The display design used TCAS-like information for all traffic and the target, where an indicated airspeed indication was added to every traffic symbol. This enables pilots to assess what speeds can be expected, thus making an estimation of the nominal speed profile. Knowledge of target's current speed and the nominal profile, which is flown by the first aircraft in the chain, is expected to yield better performance.

Spacing Marker and Allowed Error Margin

A *spacing marker* indicated the position along-track where own should be. To rule out exact error tracking, a spacing requirement *area* or *spacing error margin* is presented to the pilot instead of the exact spacing requirement. The allowed error was 5% of the requirement. It is hypothesized that if pilots are allowed to have some spacing error, not every speed change by the target aircraft is followed. In this way speed errors made by preceding aircraft are expected to be "filtered out", improving chain stability.

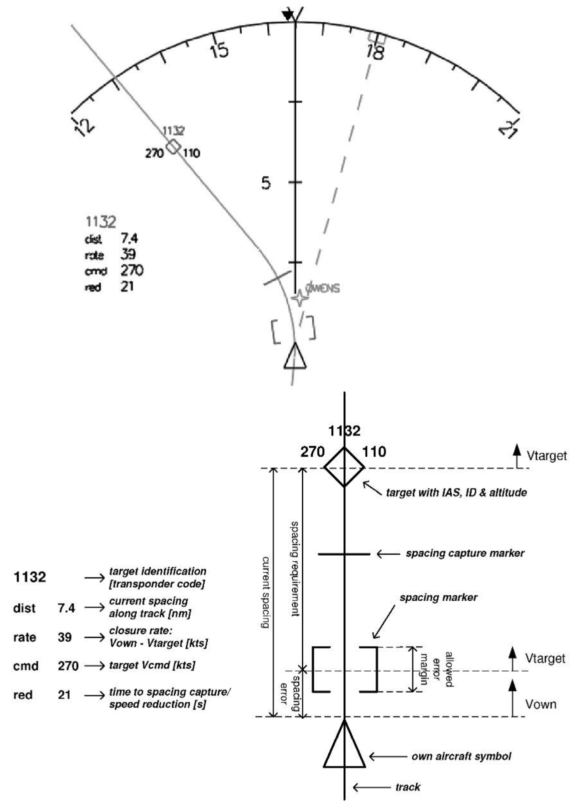


Figure 2 The Boeing 747-400 Navigation Display, augmented with self-spacing symbology.

Spacing Capture Marker

The *spacing capture marker* (SCM) can assist pilots during spacing reduction. Here the spacing error reduces since target is flying at a lower speed then own. At some point in the future the error will be zero and the spacing requirement will be met. The SCM marks the location along-track where the spacing requirement will be met, and calculates the time to get to this location. It takes into account the time needed to decelerate to the target's speed and achieve a zero closure rate. The SCM uses a linear deceleration model that predicts the very moment at which the pilot should reduce speed (event 4 in Figure 1). When the marker is reached, pilots can reduce and select a speed that matches the ground speed of the target.

Experiment

Subjects and Instructions to Subjects. Twenty professional airline pilots participated in the experiment. The first four pilots each did twelve experiment runs, where the remaining sixteen pilots did sixteen runs. Pilots were introduced to the general concept of self-spacing, and more specifically to the principles of the procedure developed here. They were instructed to execute the spacing reduction phase by maintaining their speed. During spacing hold, pilots were instructed to use the spacing error margin in case of deviant behavior of their targets.

Independent variables and experiment design. Three independent variables were tested.

Two procedures were defined. First, the *controller's initiative* procedure, where pilots received spacing instructions directly from the controller. The spacing instructions were tied to the sections that defined the arrival and the nominal speed profile, through waypoints. The second procedure is the *pilot's initiative* procedure, where it is the pilot's task to determine and select the correct spacing requirement in case of a speed reduction of the target aircraft. A table with the correct spacings for several speeds was shown on the arrival chart.

Four different displays were used. They incorporated all features introduced above. To assess the usability of the SCM and V_{cmd} indication, these features were not always present in the display. This results in four display configurations.

Four chain positions, namely positions 2 to 5, were used. The aircraft flying at position 1 was pre-recorded and followed a perfect nominal speed profile. Every run was recorded and played back: a pilot would be following a target aircraft that was actually flown by a previous experiment pilot.

The three independent variables yield 32 experiment conditions. These would require 32 pilots, who each fly 32 runs. This amount of runs would require too much time, and the amount of combinations and therefore pilots is cut in half. Pilots still fly each combination of 'procedure' and 'display' but only half of the possible combinations of 'procedure and display' and 'position'. The remaining sixteen combinations included a different set of positions for each pilot, while still all four positions would be flown four times by each pilot.

Arrival scenarios. Nine arrival scenarios were defined, where each scenario *arrival route* shared the same underlying structure (Figure 3). Each route would be rotated, mirrored and given an altitude offset yielding a different scenario for each run.

In section 2 a disturbance is introduced in the scenario. In section 2A the speed of 300 IAS and a nominal altitude of 12000ft, together with a time spacing of 81 seconds dictates a spacing of 8.0 nm. When aircraft entered section 2B pilots had to descend 1000ft. While still flying at 300 IAS, the lower altitude causes the true airspeed, and hence the ground speed, to drop a few knots. Trailing aircraft, still flying 1000ft higher and trying to maintain a low closure rate, would be forced to slow down a little.

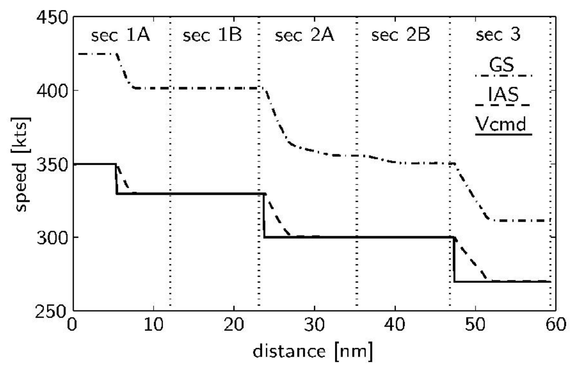


Figure 3. *Nominal speed profile where speed reductions are tied to the sections of the arrival route.*

Apparatus. The experiment was conducted in a fixed base simulator. This simulator included two 18" LCD screens on which a Primary Flight Display, Navigation Display and virtual Mode Control Panel (MCP) were shown. The autopilot was engaged during the entire run. Pilots could select autopilot speed and altitude targets via the MCP. The experiment leader acted as air traffic controller.

Aircraft and weather model. A non-linear B747 200 model was used. An ISA standard atmosphere was used and no wind was present.

Dependent measures. Since the nominal arrival would only require three speed reductions, the *number and size of speed changes* during the runs is the first measure. The second measure is the error of the *ground speed trace* of a run compared to the nominal speed profile. The error is measured over *time* since the experiment tries to separate the aircraft with a constant time-spacing equivalent. The third measure is spacing performance, i.e., the time that a pilot remains in spacing hold and the average spacing error. Workload was assessed using NASA TLX.

Hypotheses. It is hypothesized that it is possible to bring about constant-time-like self-spacing behavior using a constant-distance procedure in a stepped speed profile, canceling any slow-down effects. It is expected that intent information like target V_{cmd} and in particular the SCM will increase performance and reduces workload. Finally it is hypothesized that pilots will be able to cancel out “chain-effects”. A chain-effect is defined by the passing on and amplification towards the back of the chain of “deviant” behavior of aircraft at the front of a chain. It will be assessed by determining how the three dependent measures speed changes, speed error and spacing error of aircraft at the back of the chain is effected by deviant behavior for the same three measures by aircraft in the front of the chain.

Results and Discussion

Number of speed changes

Where the nominal speed profile only required three speed reductions the average number of speed changes was slightly more than 6. It was not affected by the procedure or the display, it slightly increased for positions further back in the chain, but this effect was not significant. This indicates that pilots were quite able to ‘filter out’ any unnecessary speed changes of aircraft flying in front of them.

A significant effect on the number of speed changes is found for the *section* in which the aircraft was flying (F=49.746, p<0.01). The number of speed changes in section 2 was almost twice the number in sections one and three. In section 2 the altitude dropped 1000ft when aircraft entered section 2B. When the *target* aircraft entered section 2B, the altitude drop causes the ground speed to reduce a little. This had to be compensated by *own*, still flying in section 2A, requiring some small speed reductions.

Overall, pilots understood the difference in tasks between the spacing hold and reduction phases. During spacing reduction pilots were to remain their current speed until the spacing requirement was met. Therefore the number of speed changes is lower as compared to the spacing hold phase. The average of about 6 speed changes are almost all accounted for during the spacing hold phase, which lasted about 70 % of the total runtime. Thus the comparison of the total amount of speed changes during spacing hold and reduction should be done with care.

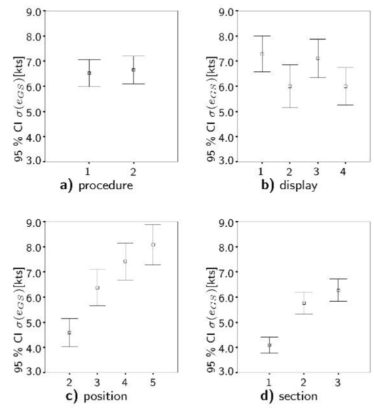


Figure 4. *Speed error with respect to nominal.*

Speed Error

Figure 4 shows the (standard deviation of) the ground speed error e_{GS}, where the ground speed of every run is compared to the *nominal speed profile*, see Figure 3. The procedure does not impose effects on the speed error. Performance significantly improved for displays with the SCM (2 and 4) (F=11.950, p<0.01).

The effect of the chain position is also significant (F=19.038, p<0.01), resulting in a growing speed error for positions further back in the chain. In sections 2 and 3 the speed error is larger then in section 1. This is caused by a slow-down effect in section 2, where the average speed decreases for positions further back in the chain. This coincides with findings for the speed changes in section 2, discussed above. The slow-down is *compensated for*, however, in section 3, where speed error grows more positive for positions further back in the chain.

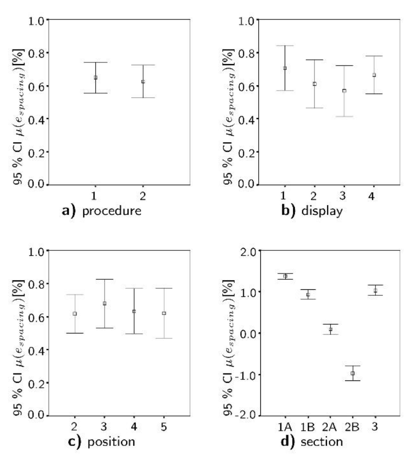


Figure 5. *Spacing error with respect to nominal.*

Spacing Performance

The spacing error with respect to the *nominal spacing profile*, i.e., tied to the arrival structure, is studied. By analyzing spacing performance with respect to the nominal spacing profile the spacing performance is assessed from a controller's perspective. Figure 5 reveals no effects of the experiment conditions on the average spacing error. However, over the *sections* the error varied much. In section 2 pilots started flying 'too-close'. Apparently the slow-down effect is compromised by pilots letting the spacing error become more negative instead of trying to fly a zero closure rate and a zero spacing error. This was substantiated by analysis of the closure rate for small spacing errors; in section 2 the average closure rate for small spacing errors is two times higher then in sections 1 and 3 (de Groot, 2004).

Chain Effects

Some chain effects are found for the speed error but these effects are not very strong. This means that high speed errors of aircraft flying at the front of the chain did not result in much higher speed errors at the back of the chain. No chain effects are present for the spacing error either, i.e., large spacing errors of aircraft flying at the front of the chain are not passed through towards the back of the chain.

Workload

The workload of the task was rated very low. No effects were found of the procedures. Displays with the SCM reduced workload, (F=3.5934, *p*=0.059). Borderline significance was found for the position in chain (F=2.358, *p*=0.072), indicating that the workload was a little higher at the back of the chain.

Subjective Comments

Pilots rated the procedure as providing enough information to assist them in the self-spacing task. They preferred the controller's initiative procedure, as they believe the controller has a better overview of the situation and therefore should remain in control of determining and issuing spacing requirements. Also time pressure is rated lower compared to the pilot's initiative procedure. Pilots noted that the latter could become a very efficient procedure if they would be allowed to follow their own vertical trajectory, with only spacing requirements at certain positions along the arrival. Finally, pilots commented that spacing requirements could be tied to waypoints instead of arrival sections or target speeds.

Displays

Generally the displays were rated as providing an appropriate level of information to the pilot. Ratings for the various display features reveal that the distance to the target aircraft and the closure rate were very helpful. Pilots commented that the spacing marker is required but should not be placed on the track because it requires a small display range. Instead they would prefer a spacing indication in the form of a bar that does not depend on the ND range.

The target V_{cmd} indication was considered superfluous since the nominal speed profile was clear and pilots did not expect the target aircraft to select off-nominal speeds. However, in off-nominal situations they indicated that V_{cmd} could be useful.

The SCM was considered very helpful, as it was reported to take away time pressure during spacing reduction. Instead of having to scan the display continuously, the SCM instantly indicates if action (speed reduction) is already required. Many pilots reported an intensive use of the speed bug attached to the target aircraft symbol. They used the speed bugs on other traffic flying down the arrival to assess what speeds are to be expected, thus creating a mental picture of the nominal speed profile.

Conclusions

The spacing reduction concept can rule out slow-down effects with distance-based spacing in approaches where speeds gradually reduce. The selection of spacing requirement by the pilot instead of the controller does not bring about more off-nominal speed and spacing behavior. Pilots commented that a procedure where speed and spacing requirements are published on arrival charts would create a workable situation. However, they noted that the controller should remain responsible and intervene in cases of off-nominal behavior of target.

No strong chain effects were found. The allowance of spacing error introduced a dampening effect because it allowed pilots more time to assess, and act to, the actions of the target aircraft. The knowledge of the nominal speed profile, provided by speed tags on traffic symbols, also retained pilots from following off-nominal behavior of target aircraft. Pilots rated the workload of the self-spacing task as very low and they believed that introducing the task into the arrival phase of the flight is possible.

References

Abeloos, A.L.M., Van der Hilst, J.A.T., Mulder, M., Van Paassen, M. M., & Hoffman, E. (2001). Design of a CDTI for Longitudinal Station Keeping. *Proceedings of the 20th Digital Avionics Systems Conference*, Daytona Beach (FL), U.S.A., October 18-20, 7.B.4/1-7.B.4/10.

de Groot, J.M.C. (2004). *Distance-Based Self-Spacing in Arrival Streams*. Unpublished MSc report, Faculty of Aerospace Engineering, Delft University of Technology, Delft, the Netherlands.

Hoffman, E., Zeghal, K., Cloerec, A., Grimaud, I., & Nicolaon, J. P. (1999). Operational concepts for limited delegation of separation assurance to the cockpit. *Proceedings of the AIAA Guidance, Navigation, Control Conference*, Portland (OR), August 9-11, AIAA paper 1999-3990.

Lee, P.U. et al., (2003). Free maneuvering, trajectory Negotiation, and Self-Spacing concepts in Distributed Air-Ground Traffic Management. *Proceedings of the 5th USA/Europe Air Traffic Management Research and Development Seminar*, Budapest, Hungary, 1–10.

Pritchett, A. R. & Jankosky, L. J. (2003). Pilot-Performed In-Trail Spacing and Merging: An Experimental Study. *Journal of Guidance, Control, and Dynamics*, 26(1), 143-150.

Sorensen, J. A. and Goka, T. (1983). Analysis of In-Trail Following Dynamics of CDTI-Equipped Aircraft. *Journal of Guidance, 6* (3), 162–169.

Williams, D. H. (1983). Self-Separation In Terminal Areas Using CDTI. *Proceedings of the Human Factors Society, Vol. 27*.

EVALUATION OF WORKLOAD IN HIGH COMPLEXITY WORK PLACE: AN EXPERIMENT DURING A REAL SITUATION

Carlos Gomes de Oliveira, D Sc
ICAF/COMAER/OACI-MCT
Afonso de Taunay 360/101, Barra da Tijuca, Rio de Janeiro, RJ, ZC: 22431-000, Brazil
Tel:fax: 55 21 21088905
oliveiracg@yahoo.com
http://www.nuicaf.aer.mil.br

Selma Leal de Oliveira Ribeiro, D Sc
ICAF/COMAER/OACI-MCT
Rua Eng Hilton Gadret, 80, Bl. 05, Apt, 105. Irajá, Rio de Janeiro/RJ - Brazil
Tel:fax: 55 21 21088942
sribeiro@nuicaf.aer.mil.br
http://www.nuicaf.aer.mil.br

Some workplace have been widely changed with regard to their automation process, which has promoted a more complex environment concerning the task performance, demanding to the operator the introducing of new abilities. In the aeronautic activity the workload also has been diversified, as the mental demand has been enhanced. The needs of determining the impact of the workload on the operator due to such work place, evidencing a more complex nature, shows to be more important, mainly when looking at the certification requirements for new aircraft development. Such certification process is responsible for determining the minimum aircrew necessary, based on the distribution of the cabin workload, as well as keeping the situation awareness during the different phases of the flight. This study uses psychological and physiological methods of measurements to evaluate the workload in real situation during the end of the certification process of an aircraft, aiming at to identify potential methods to be implemented during the whole certification process. A protocol of workload evaluation was implemented based on the use of interview, NASA-TLX scale, heart rate (HR) and heart rate variability (HRV)**.** Two pilots participated in the study. The measurements and interviews were conducted during flights performed in the final certification process of an aircraft produced in Brazil. A total of six take-off and six landings performed during three consecutive days were evaluated. Each route was previously determined, which involved some abnormal situations according to an established program for the evaluation of the aircraft in terms of human factor requirements. The data analysis was performed in a descriptive and qualitative basis due to the peculiarity of each task. Preliminary results indicate the landing to be more stressful than take-off, and for such situations, the pilot flying (PF) had the more workload during the tasks than the pilot monitoring (PM)**.** When comparing all flights and their tasks, no important difference between the HR and HRV was observed, but, again, the landing showed a little higher stressful than take-off for the PF, as evidenced by the HR. However, the general results, including those from NASA-TLX, suggested a low workload for all tasks. With regards to the interviews, the more pronounced mental demands reported by the pilots in managing any fault of the aircraft were in those tasks that required anticipation, attention and monitoring procedures. Future studies should be conducted with the whole certification process and other scenarios in order to test the applicability of the methodology employed in the present study.

Introduction

Automation in aviation has promoted an increase in the complexity of the task performance of pilots due to the technological development. This automation has been introduced to increase the aircrew wellness, and, mainly, to minimize accidents, given that it has reduced the human error responsible for about 70% of accidents and incidents in aviation (BILLINGS, 1997). It is remarkable that the pilot's cockpit has had one of the most significant improvements aiming at the workload reduction, due to the automated devices, mainly in terms of releasing the physical workload of aircrew. However, the modifications performed so far have changed the workload of the aircraft operator, and its mental component has been enhanced while the physical has decreased.

The certification of new aircraft in term of human factor aspects has been applied, aiming at determining the workload of such aircraft, and minimum aircrew. The requirement of the establishment of minimum aircrew looks at a better distribution of workload during the different phases of flight (WISE & WISE, 2000; TATTERSALL, 2000). It is necessary to maintain a balance between demand of tasks and the capacity of the operator with different objectives, including those required to evaluate items related to certification of new aircraft for human factors. The literature shows a consistent search for assessment of mental workload by the use of subjective and physiological methods (BACKS, 1995). The main problem arises when one intends to measure the workload of pilots in cockpits, and to establish its minimum and/or maximum level permitted.

Objecting the further use in aircraft certification, RIBEIRO & de OLIVEIRA (2003) proposed a method for evaluation overall workload in pilots, which was firstly experimented in simulated flights and showed to be useful.

The present study evaluates the workload during real flights conducted during the last phase of the certification process of an aircraft aiming at identify potential methods of evaluation workload in such process.

Methods

The study was conducted during the certification process of an aircraft made in Brazil. Due to the complexity of the experimental protocol and availability of flights, only two high experienced pilots were monitored. They alternated the position of pilot flying (PF) and pilot monitoring (PM), but not in the same flight. Six flights were monitored during three consecutive days. Two phases were evaluated, take-off (began when the engine one was switched on and finished when the aircraft reached 15,000 ft), and landing (began at 10,000 ft and ended when all engines were off). The team formed by the certification authority, and the manufacturer technical staff determined each route and abnormal situations that occurred during the flight, considering the aircraft evaluation in terms of human factors. The research group did not take part in this process. The abnormal situations included the absence of electric, hydraulic and other automated systems during the flight.

Instruments of Evaluation

When compared to physical, the mental or cognitive workload is considered a little more difficult to be assessed (KANTOWITZ & CASPER, 1988). Combining the use of physiological and subjective techniques is more recommended, and has been considered as a better prediction of the workload in tasks of systems in development or implementation, with less interference in the task (WIERWILLE & EGGEMEIER, 1993). Thus, physiological and subjective techniques were employed in this study.

Physiologic evaluation: Heart Rate (HR) has been applied as a measurement of workload. Additionally, power spectral analysis of Heart Rate Variability (HRV) is a sensitive index of autonomic activities. Within the HRV, two main components have been identified, the Low Frequency (LF) at 0.03-0.15 Hz, reflecting both sympathetic and parasympathetic activity, and High Frequency (HF) at 0.15-0.4 Hz, which reflect the parasympathetic tone of the sinusoidal respiratory arrhythmia. LF/HF ratio has been proposed as an index that reflects the balance of the autonomic nervous activity (TASK FORCE of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996). Moreover, previous studies have revealed a relationship between sympathetic activity and mental effort (SATO et al., 1998; KAMADA et al. 1992). Thus, in the present research the physiological evaluation was performed through the measurement of the HR and the analysis of the HRV.

The electrocardiogram (ECG) signal was captured and simultaneously digitally recorded in a ME3000P8 (Mega Electronics), after sampled at 1000 Hz. A specific program to detect the R-waves of the ECG signal and construct the RR intervals was developed in Matlab 5.02c (Mathworks). The time series formed by the RR intervals were thus interpolated so as the sample rate of the respective HRV signal was 2 Hz. The Heart Rate (HR) was calculated as the inverse of the mean of HRV. The power spectral was estimated through Auto Regressive model with an order of 12. From the HRV signal, the power of the LF band (between 0.03 and 0.15 Hz), the power of the HF band (between 0.15 and 0.4 Hz) was determined, and LF/HF computed. Prior to each flight the ECG of the pilots were registered during a rest period of 4 minutes. HR and LF/HF, determined in each phase of the flights, were further normalized with respect to those respective variables computed during the rest test.

Subjective evaluation: To evaluate mental workload the subjective techniques are more often applied. In such case, the perception of the worker to his performance in a specific task is used. This can be considered as indices of global sensitivity to the workload (WIERWILLE & EGGEMEIER, 1993). The subjective technique performed in this study is the Task Load Index Scale - NASA – TLX (HART & STAVELAND, 1988), considering their sensitivity which has showed to be consistent in many studies with different levels of demand (HARRIS *et al.*, 1995; HANCOCK *et al.*, 1995). The TLX has six components to measure workload: mental demand, physical demand, temporal demand, performance, effort and frustration level. The test was applied after the end of each phase of the flight. It was also included a sheet with the registers of activity/time during flights and interviews. After each flight a general interview was conducted with each pilot regarding workload, automation and performance.

Results

With regards to the NASA-TLX, only the results of physical demand (PD) and mental demand (MD) components of workload will be presented. Flights are numbered from 1 to 6, and some data from the flight number 6 were missed. When a pilot was in the PF position in the forward direction (A) of the route, during take-off (T) and landing (L), the other one was PF in the backward direction (B) of the same route.

When in the PF position, P1 presented higher PD and MD during most landing than take-off (Figure 1), but when assuming the PM position, the MD and PD did not show this behavior, alternating in intensity during T and L, independently of type of flight (Figure 2).

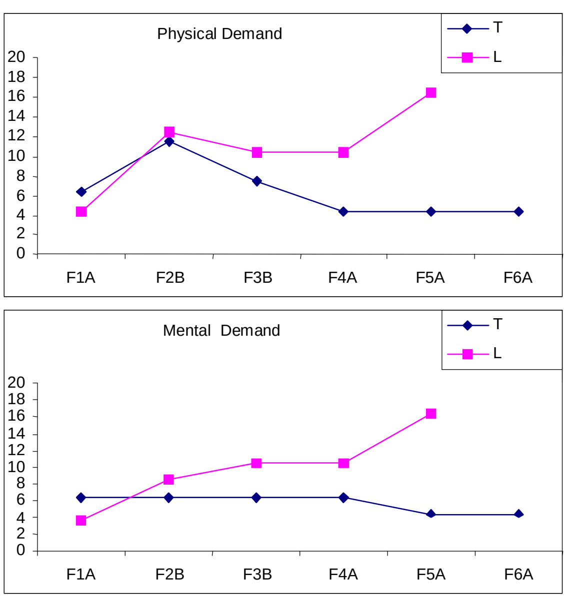


Figure 1. *Results of NASA-TLX of P1 as PF.*

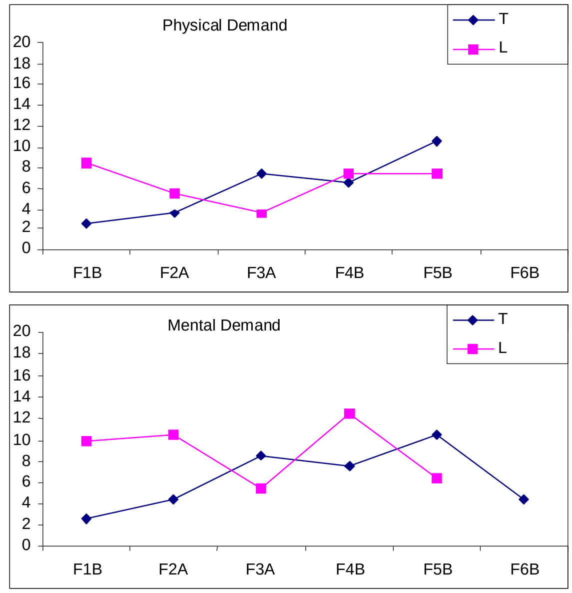


Figure 2. *Results of NASA-TLX of P1 as PM.*

The pilot P2 did not follow the same behavior. In the PF position, P2 showed PD to have almost the same value during take-off and landing (Figure 3). This was also true for this pilot while in the PM position (Figure 4). These results indicate that independently of the flight, which was related to different abnormal situation, no difference in the workload was perceived by this pilot among all performed tasks.

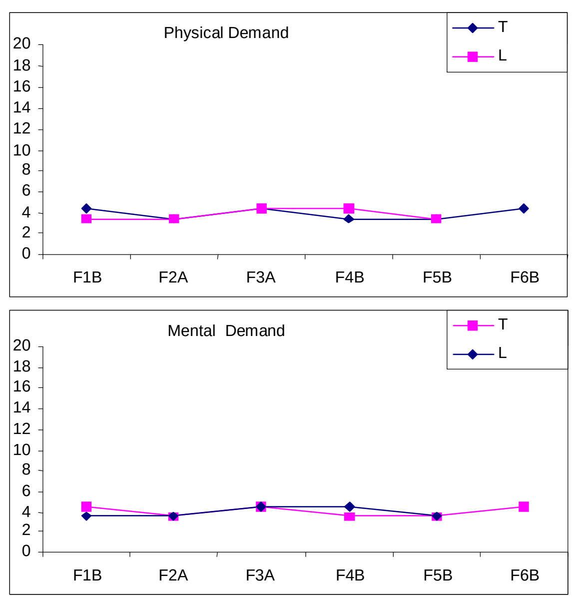


Figure 3. *Results of NASA-TLX of P2 as PF.*

The physiological measurements corroborated the results of many findings of the NASA-TLX in some aspects. The HR is expressed as percentage above that found during the rest test, and the HRV as the ratio between the values of the flight and the rest.

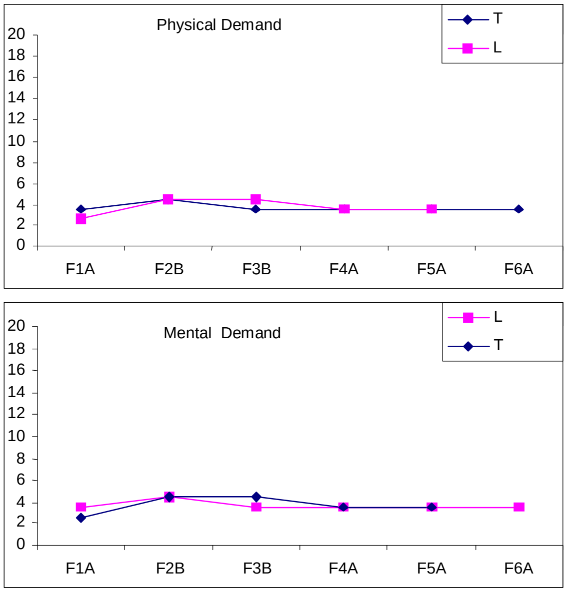


Figure 4. *Results of NASA-TLX of P2 as PM.*

The HR of pilot P1 as in the PF position was higher during landing than during take-off in all flights (Figure 5). The result of HRV also showed the LF/HF higher during landing than take-off (Figure 5), suggesting higher mental workload during landing. On the other hand, as PM, no clear pattern was observed for HR or HRV in this pilot (Figure 6).

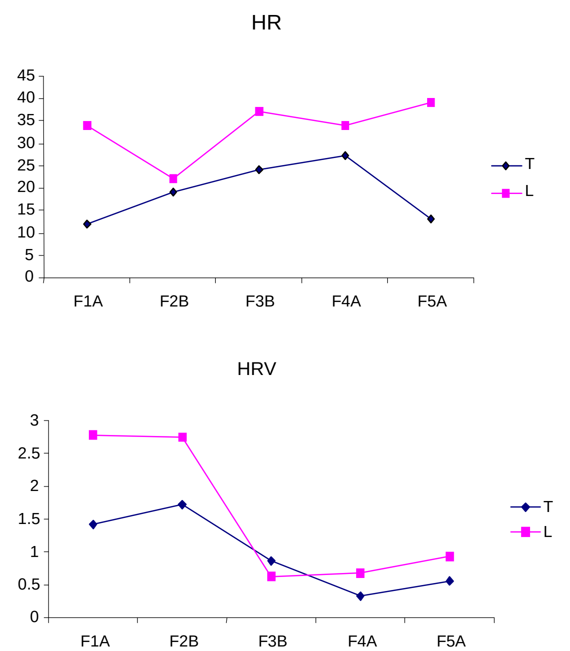


Figure 5. *Results of HR and HRV of P1 as PF.*

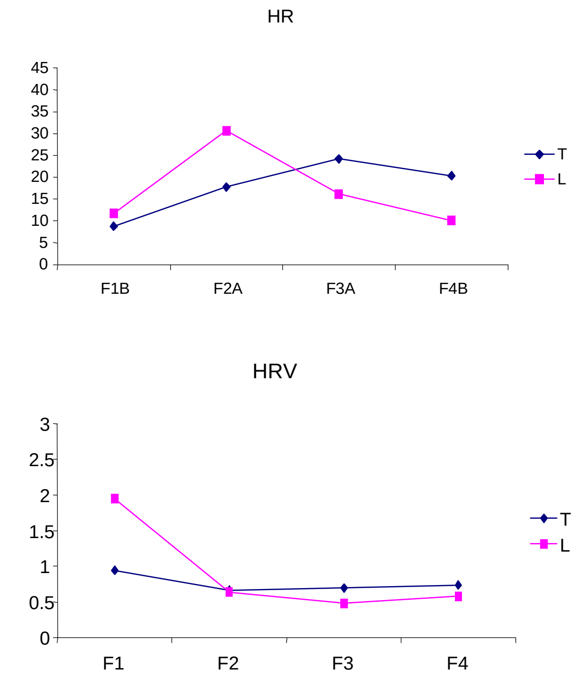


Figure 6. *Results of HR and HRV of P1 as PM.*

With regards to pilot P2 in the PF position, HR was higher during landing than take-off in all flights and little difference was observed in the HRV when comparing the phases of flight (Figure 7). When in the PM position, again no pattern was observed for HR and HRV, and in one flight the HR was lower than that presented during the rest test (Figure 8).

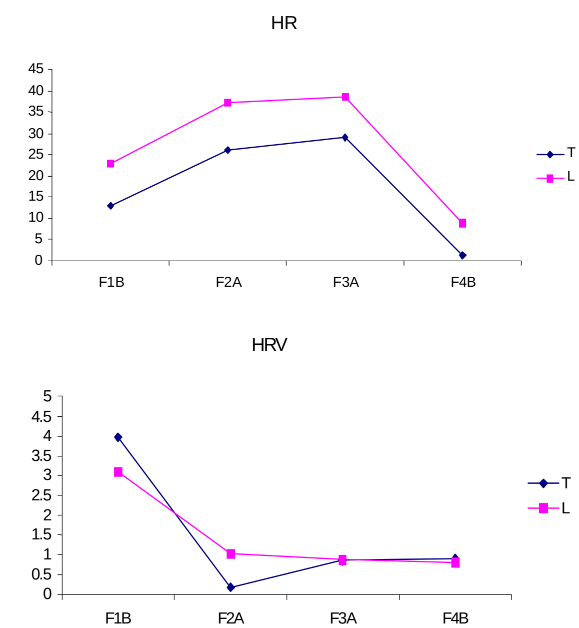


Figure 7. *Results of HR and HRV of P2 as PF.*

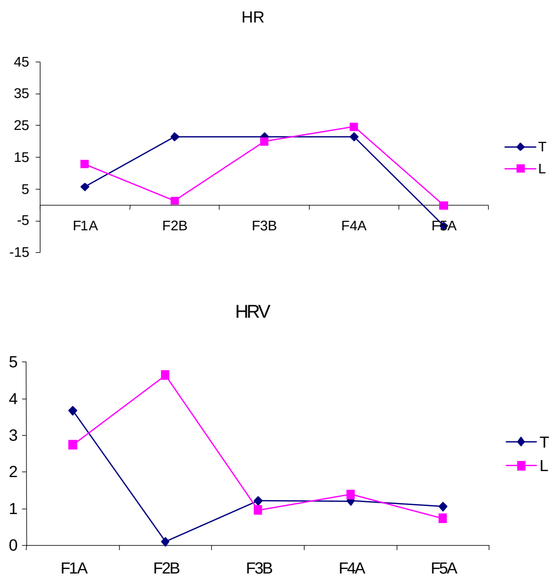


Figure 8. *Results of HR and HRV of P2 as PM.*

Concerning the interviews, during the most time evaluated the pilots did not report important physical demands due to the abnormal situations in the flight. The most relevant report that should be pointed out is the mental demand due to the anticipation. This was attributed to the intrinsic characteristic of the situation caused by the necessity of pilot's anticipation.

Discussion

The international institutions of regulation in aviation have proposed the certification process of aircraft for human factors. Thus, specific methods of measurements of workload have to be developed. This work investigated the use of some methods in the evaluation of pilot's workload only during a short window of the last part of the certification process of a new aircraft.

Although the study evaluated only two pilots, when comparing the workload of PF and PM, there is no clear suggestion that the first has higher demand than the second, either physical or mental. This is not in agreement with the study of RIBEIRO & de OLIVEIRA (2003), who suggested higher mental demand for PM than PF and the contrary for physical demand. The difference between the aircraft used in the present work and the simulator used by RIBEIRO & de OLIVEIRA (2003) in them study might explain these controversy results. Despite the presence of abnormal condition in the flights of this investigation, the aircraft is thoroughly atomized and, thus, even during abnormal condition, could not have highly introduced extra mental or physical workload. Another reason is the high experience of the pilots with the aircraft, which was obtained during all the process of its certification.

HR showed to be potentially able to identify differences between positions and between tasks. The higher value of HR found in PF than PM is not surprising because the literature has previously reported this difference. According to BACKS (1995) HR is higher in the pilot who is in control, than in other aircrew and decreases when the pilot leaves the control and increases in the pilot taking over the control. Furthermore, it should be taken into consideration that these results are related to take-off and landing while during cruise HR might be expected do decrease.

The most interesting discussion arises when interpreting the results of HRV. VELTMAN & GAILLARD (1998) show that mental effort suppresses the activity of the cardiovascular control system, suggesting that there is more respiratory activity during rest than during a task in the LF band (<0.15 Hz) in mental tasks, which thus make difficult to interpret the effect of mental workload in HRV. In the present study the HRV was evaluated by means of LF/HF ratio since studies have proposed that during mental effort this ratio tend to increase when compared to the rest (SATO et al., 1998; KAMADA et al. 1992). As presented in the results of NASA-TLX, the HRV suggested more mental workload during landing than take-off and no clear difference could be observed concerning the different position assumed by the pilots.

It has been hypothesized that autonomic responses such as of HR are multidimensional determined and not just reciprocally coupled, meaning that there might be an activation of one branch with the inhibition of the other or even the co-activation and co-inhibition. Thus, although the HRV has been showed as a potential tool to evaluate autonomic response, even in the present study, their results should be interpreted carefully.

An important finding concerning the physiological measurements is the consistency of the data when focusing a pilot in particular. The values found are within a short range, and short range was also observed in the results of NASA-TLX. In general, the workload appeared to be low and little difference were observed when comparing the different flights. One question that still remains is how to quantify the workload in an objective criterion.

With regards to the item stressed by the pilots during the interview - the anticipation -, this is expected to be present when the pilots have to analyze the possible consequences of any atypical situation that the aircraft has to be submitted to. They have to do this anticipation in order to be prepared for other unexpected abnormal occurrence, as bad meteorological and/or visibility condition. During landing, some of the abnormal conditions involving the suppressing of automation devices, as the electric fail in the flight F5A, were described as demanding from the PF abilities required in traditional flights, since the automation was not available.

Conclusion

As a conclusion, the methods employed appear to constitute in a good tool in the evaluation of workload during the certification process of aircraft for human factors. The main problem that still remains is to establish the minimum and maximum values for the variables measured in order to define what is the desired or undesired workload when certifying a new aircraft. It should be also taken into consideration that this study investigated the use of some methods in the evaluation of pilot's workload during the last part of the certification process of a new aircraft. In fact, the study should be extended to the whole process and should have the participation of more pilots.

Acknowledgments

This study was developed with the support of the Brazilian Civil Aviation Department through the Technical Cooperation Mission of the International Civil Aviation Organization and the technical cooperation of the Aeronautical Certification Division.

References

BACKS, R. W. (1995) Going beyond heart rate: autonomic space and cardiovascular assessment of mental workload. In: *The International Journal of Aviation Psychology*, 5 (1), 25-48.

BILLINGS, C. E. (1997). *Aviation automation: the search for a human-centered approach.* New Jersey: Lawrence Erlbaum Associates, Publishers.

GAILLARD, A. W. K. Comparing the concepts of mental load and stress. *Ergonomics*, 1993, vol. 36, n. 9, p. 991-1005

HANCOCK, P. A., WILLIAMS, G., MANNING, C. M. (1995) Influence of task demand characteristics on workload and performance. In: *The International Journal of Aviation Psychology*, 5 (1), 63-86.

HARRIS, W. C., HANCOCK, P. A., ARTHUR, E. J., CAIRD, J. K. (1995) Performance, workload, and fatigue changes associated with automation. In: *The International Journal of Aviation Psychology*, 5 (2), 169-185.

HART, S. G. & STAVELAND, L. E. (1988) Develpment of NASA-TLX (task Load Index): results of empirical and theoretical research. In: HANCOCK, P. A. & MESHKATI, N. (Eds) *Human Mental Worload*. Amsterdam: Elsevier.

KAMADA T, KUMASHIRO, M; MONOU, H; INOUE, K; Power spectral analysis of heart rate variability in Type As and Type Bs during mental workload. *Psycosom Med*, 1992, 54, 462-470

KANTOWITZ B.H. and CASPER, P. A. (1988) Human workload in aviation. In WIENER, E. L. & NAGEL, D. C. (Eds.) *Human factors in aviation.* Academic Press: San Diego, California.

RIBEIRO, S.L.O and de OLIVEIRA, C.G. (2003) Workload in flight cockpit: an approach for searching a methodological evaluation, Proceedings of the International Symposium on Aviation Psychology, Ohio, USA.

SATO, N.; KAMADA, T.; MIYAKE, S.; ARAKATSU, J.; KUMASHIRO, M.; KUME, Y. (1998) Power spectral analysis of heart rate variability in type A females during a psychomotor task, *Journal of Psychomotor Research.* 45(2), 159-169.

TASK FORCE of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. “Heart Rate Variability: Standards of Measurement, Physiological Interpretation and Clinical Use”, *Circulation* 93(5):1043-1065, Mar/1996.

TATTERSALL, A. J. (2000) Practical guidelines for workload assessment. In: WISE, J. A. and HOPKIN, V. D. (Eds.). *Human Factors in Certification.* New Jersey: Lawrence Erlbaum Associates, Publishers.

VELTMAN, H. & GAILLARD, A. (1998) Physiological workload reactions to increasing levels of task difficulty, *Ergonomics*, 1998, 41(5), 565-669

WISE, M. A. & WISE, J. A. (2000) The use of the systems approach to certify advanced aviation technologies. In: WISE, J. A. & HOPKIN, V. D. (Eds.). *Human Factors in Certification.* New Jersey: LEA.

LAND OF THE MIDNIGHT SUN: SHEDDING LIGHT ON DIFFERENCES IN GA ACCIDENTS IN ALASKA VERSUS THE REST OF THE UNITED STATES

Cristy Detwiler,[1] **Carla Hackworth**,[1] **Kali Holcomb**,[1] **Albert Boquet**,[2]
Elaine Pfleiderer,[1] **Douglas Wiegmann**,[3] and **Scott Shappell**[4]
Federal Aviation Administration[1], Embry-Riddle Aeronautical University[2], Mayo Clinic[3], Clemson University[4]

General aviation (GA) accidents that occurred in Alaska versus the rest of the United States were compared using the Human Factors Analysis and Classification System (HFACS). Overall, categorical differences among unsafe acts (decision errors, skill-based errors, perceptual errors, and violations) committed by pilots involved in accidents in Alaska and those in the rest of the U.S. were minimal. However, a closer inspection of the data revealed notable variations in the specific forms of unsafe acts within the accident record. Specifically, skill-based errors associated with loss of directional control were more likely to occur in Alaska than the rest of the U.S. Likewise, the decision to utilize unsuitable terrain was more likely to occur in Alaska. Additionally, accidents in Alaska were associated with violations concerning VFR into IMC. These data provide valuable information for those government and civilian programs tasked with improving GA safety in Alaska and the rest of the US.

Introduction

Considerable effort has been expended over the last several decades to improve safety in both military and commercial aviation. Even though many people have died and millions of dollars in assets have been lost, the numbers pale in comparison to those suffered every year within general aviation (GA). For example, according to the National Transportation Safety Board (NTSB), there were 1,741 GA accidents in 2003 that resulted in 629 fatalities (NTSB, 2005). While the numbers may not register with some, when considered within the context of commercial aviation, the losses suffered annually by GA are roughly equivalent to the complete loss of three commercial passenger Boeing 727's.

Why then has GA historically received less attention? Perhaps it has something to do with the fact that flying has become relatively common as literally millions of travelers board commercial aircraft daily to get from place-to-place. Not surprisingly then, when a commercial airliner crashes, it instantly becomes headline news, shaking the confidence of the flying public.

In contrast, GA accidents happen virtually every day yet they receive little attention and seldom appear on the front page of *USA Today*. Perhaps this is because they happen in isolated places, involving only a couple of unfortunate souls at a time. In fact, unless the plane crashed into a school, church, or some other public venue, it is unlikely that anyone outside the local media, government, or those intimately involved with the accident even knew it happened.

Over the last couple of years, GA has deservedly received increasing attention from the FAA (FAA Flight Plan 2004-2008) and other safety professionals. Indeed, several groups from the government (e.g., the FAA's Civil Aerospace Medical Institute; National Institute of Occupational Safety and Health), private sector (e.g., the Medallion Foundation), and universities (e.g., University of Illinois, Johns-Hopkins University) have conducted a number of studies examining GA accident causation.

Alaskan Aviation

It is of note that many of these efforts have focused on Alaska, where aviation is the primary mode of transportation. Alaska is known for its varied and often unique landscape and when this is considered with temperamental weather and seasonal lighting conditions, even the most experienced pilot would have to agree that Alaskan aviation represents some of the most difficult flying in the U.S., if not the world. The combination of factors mentioned above, the number of GA accidents that are occurring in Alaska and the FAA's accident reduction goal (FAA Flight Plan 2004-2008) were factors in our decision to implement this study.

Human Error and General Aviation

A variety of studies have been conducted in an attempt to understand the causes of GA accidents. Most have focused on contextual factors or pilot demographics, rather than the underlying causes of the accidents. When the leading cause of accidents, human error, has been addressed, it is often only to report the percentage of accidents associated with aircrew error in general or to identify those where alcohol or drug use occurred. What is needed is a thorough human error analysis. Previous attempts to do just that have met with limited success (O'Hare, Wiggins, Batt, & Morrison, 1994; Wiegmann & Shappell, 1997). This is primarily because human error is influenced by a variety of factors that are usually not addressed by traditional classification schemes (Shappell & Wiegmann, 1997). Yet, with the development of the Human Factors Analysis and Classification System (HFACS) previously unknown patterns of human error in aviation accidents have been uncovered (Shappell & Wiegmann, 2001; Wiegmann & Shappell, 2001a).

Method

GA accident data from calendar years 1990-2002 were obtained from databases maintained by the NTSB and the FAA's National Aviation Safety Data Analysis Center (NASDAC). In total, 24,978 GA accidents were extracted for analysis. Only accidents occurring during 14 CFR Part 91 operations were included (22,987 cases). This analysis was primarily concerned with powered aircraft and thus the data were further restricted to include only accidents involving powered fixed-wing aircraft, helicopters, and gyrocopters. The remaining 22,248 accidents were then examined for aircrew-related causal factors. In the end, 17,808 accidents were included in the database that were associated with some form of human error and submitted to further analyses using the HFACS framework.

Results

When using HFACS to examine the GA accident data, the majority of the accidents are coded with either a precondition for unsafe acts or an unsafe act. This is due primarily to the fact that there is typically not much of an organizational structure or supervisory influence on the majority of GA pilots, as compared to their counterparts conducting commercial or "for hire" operations.

Indeed, with few exceptions (e.g., flight instructors and flight training institutions), the top two tiers of HFACS (unsafe supervision and organizational influences) remained sparsely populated when examining the GA accidents leaving the majority of causal factors within the bottom two tiers of HFACS. Consequently, the balance of this report will focus only on the unsafe acts of the operator level of the HFACS framework.

Unsafe Acts of Operators (Aircrew)

An overall review of the GA accident data yielded the following results (see Figure 1). The most prevalent error noted in the accident data over the past decade was skill-based errors (73%), followed by decision errors (28%), violations (13%), and perceptual errors (7%).[1] The relatively flat lines in the types of unsafe acts across the years suggest that past intervention strategies have had little differential impact on any particular category of error.

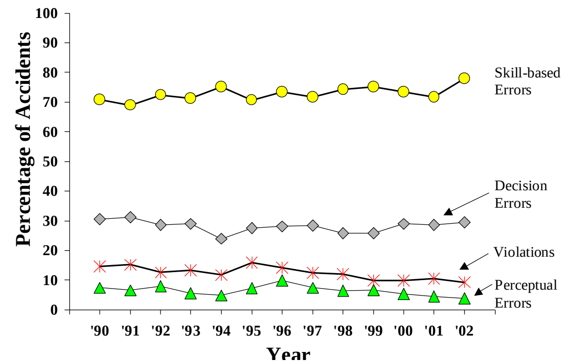


Figure 1. *Overall review of general aviation data for HFACS unsafe acts.*

To obtain a better sense of how human error differences between Alaska and the rest of the United States (RoUS) are represented in the data, the error types were broken out accordingly (Figure 2). The analysis of the unsafe acts revealed that there were slightly more decision errors, fewer skill-based errors, perceptual errors and violations in Alaska than there were in the RoUS.

Note, the following analyses did not distinguish between those pilots who were native to Alaska and were involved in an accident versus those who were less familiar with the state. That being said, the numbers for Alaska reflect the accidents that occurred within the physical boundaries of the state.

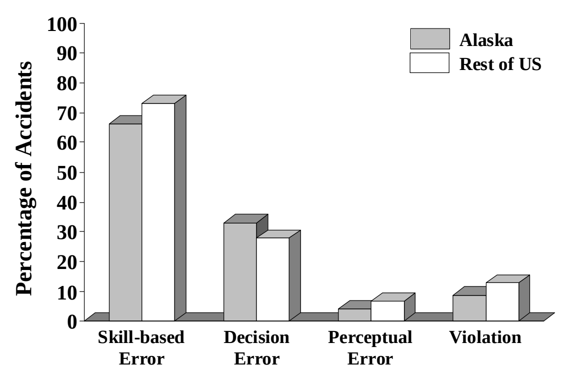


Figure 2. *Percentage of accidents associated with each of the unsafe acts of the operator.*

[1] These percentages do not add up to 100 because an accident could be assigned more than one HFACS code (i.e., DE, SBE, PE, etc..).

Skill-based Errors. Differences that existed between Alaska and the RoUS were fairly consistent across the years of study, with slightly more skill-based errors associated with accidents in the RoUS (see Figure 3). The only exception involved 1991, 1996, and again in 2002 where the percentages were nearly equal.

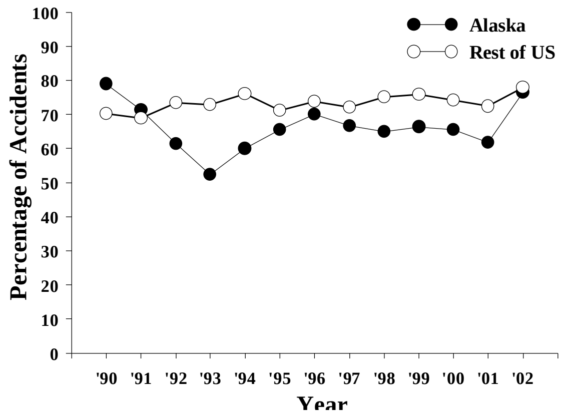


Figure 3. *Skill-based errors broken out by Alaska versus the RoUS.*

Differences between Alaska and the RoUS were more distinct when the actual type of skill-based error was compared (Table 1). Directional control was the most frequently cited skill-based error for both Alaska (19%) and for the rest of the U.S. (13%). Pilots in Alaska were more likely to experience a loss of directional control of their aircraft than those in the rest of the U.S. (odds ratio = 1.593, X^2 = 33.400, p <.001). Additionally, inadequate compensation for wind conditions was almost three times more likely to occur in Alaska, (odds ratio = 2.884, X^2 = 150.893, p <.001). Conversely, pilots in the rest of the U.S. were almost two times more likely to demonstrate airspeed errors than those in Alaska, (odds ratio = 1.733, X^2 = 20.652, p <.001).

Table 1. *Top 5 Skill-based errors occurring for Alaska and the rest of the U.S.*

Alaska	N (%)	RoUS	N (%)
Directional Control	206 (18.6%)	Directional Control	2139 (12.6%)
Compensation for Wind Conditions	170 (15.4%)	Airspeed	1932 (11.3%)
Stall	88 (8.0%)	Stall	1312 (7.7%)
Airspeed	76 (6.9%)	Aircraft Control	1310 (7.7%)
Ground Loop/Swerve	50 (4.5%)	Compensation for Wind Conditions	1009 (5.9%)

Decision Errors. To better understand the complexity of the decision errors that were occurring in the accidents for both Alaska and the rest of the U.S., a fine-grained analysis of the data was conducted. Figure 4 illustrates the decision error trends for Alaska and the rest of the U.S. across the thirteen-year period from 1990-2002. With the exception of 1990, 1991, and 2002 any difference that did exist was remarkably consistent across years of the study.

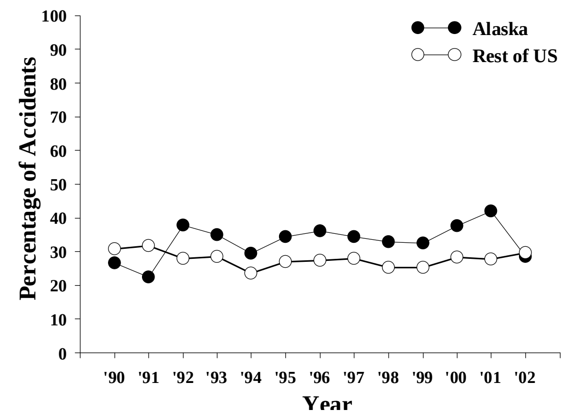


Figure 4. *Decision errors broken out by Alaska versus the rest of the U.S.*

Upon closer examination, the largest proportion of decision errors in the RoUS involved in-flight planning/decision making, accounting for 19% of those observed. However, the top decision error for pilots flying in Alaska dealt with decisions to utilize unimproved landing, takeoff, taxi areas, or unsuitable terrain. As a matter of fact, those flying in Alaska were almost 15 times more likely to takeoff and land from unsuitable terrain than those in the RoUS (odds ratio = 14.703, X^2 = 829.461, p <.001). A break-out of the top 5 decision errors for Alaska versus the rest of the U.S. is presented in Table 2.

Table 2. *Top 5 Decision errors occurring for Alaska and the RoUS.*

Alaska	N (%)	RoUS	N (%)
Unsuitable Terrain	193 (40.5%)	In-flight Planning/ Decision	1002 (18.7%)
In-flight Planning/ Decision	59 (12.4)	Planning/ Decision	374 (7.0%)
Aborted Takeoff	28 (5.9%)	Refueling	351 (6.5%)
Planning/ Decision	19 (4.0%)	Remedial Action	339 (6.3%)
Go-around	18 (3.8%)	Go-around	336 (6.3%)

Violations. In general, violations were associated with less than 20% of GA accidents (Figure 5). For the entire U.S. sample, nearly 50% of these accidents resulted in a fatality. When examining accidents in Alaska separately from the RoUS, differences were found. Accidents involving violations in Alaska were 9 times more likely to result in a fatality (odds ratio = 9.248, X^2 = 127.606, p <.001); whereas, those that occurred in the rest of the U.S. were 4 times more likely to result in a fatality, (odds ratio = 4.410, X^2 = 1054.059, p <.001).

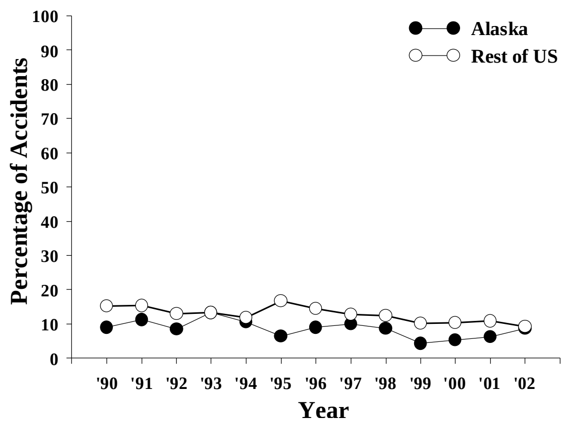


Figure 5. *Violations broken out by Alaska versus the RoUS.*

A closer look at the types of violations revealed that the most frequently cited violation for all GA accidents was Visual Flight Rules (VFR) flight into instrument meteorological conditions (IMC), (Table 3). VFR flight into IMC alone accounted for one-third of the violations in the Alaska data and was over two and a half times more likely to occur there than in the RoUS (odds ratio = 2.629, X^2 = 22.467, p <.001). Furthermore, when the weather-related violations were combined (VFR into IMC, flight into known adverse weather, and flight into adverse weather), nearly half of the violations in the Alaska data were represented.

Table 3. *Top 5 Violations occurring for Alaska and the rest of the U.S.*

Alaska	N (%)	RoUS	N (%)
VFR into IMC	38 (32.5%)	VFR into IMC	369 (15.5%)
Aircraft Weight & Balance	13 (11.1%)	Operation with Known Deficiencies	261 (10.9%)
Procedures/ Directives	12 (10.3%)	Procedures/ Directives	248 (10.4%)
Flight into Known Adverse Weather	11 (9.4%)	Flight into Known Adverse Weather	212 (8.9%)
Operation with Known Deficiencies	8 (6.8%)	Aircraft Weight & Balance	149 (6.2%)

Discussion

On the surface, there were no major differences between Alaska and the rest of the U.S. with regard to the overall pattern of human error. If anything, there were slightly more decision errors associated with accidents occurring in Alaska and fewer skill-based errors, perceptual errors, and violations. This information is similar to research in other aviation operations, which identified skill-based errors as the most commonly occurring type of error (Shappell & Wiegmann, 2003; Wiegmann & Shappell, 2001b; 2003).

The accident data suggest that aircraft handling should be taken into account when determining where interventions should be applied. For instance, any training (both *ab initio* and recurrent) along these lines should include control of the aircraft on the ground (e.g., ground loops), crosswind landings, avoiding and recovering from stalls, and general control of the aircraft in flight. Given the inherent risk associated with some of these maneuvers, it makes sense to utilize modern simulators during this training. Unfortunately, it is unclear whether there would be adequate transfer of training for these specific tasks to make simulation training viable. Therefore, before utilizing simulation to address these issues, research needs to be conducted to determine the best role simulators might play. In the meantime however, it appears necessary to emphasize these topics during actual in-flight training.

The only notable exception among the HFACS casual categories involved decision errors. Specifically, pilots in Alaska were more likely to utilize unsuitable terrain for landing, taxi, and takeoff. It would appear that educating aviators on the hazards of utilizing frozen rivers or gravel bars, for example, may reduce these types of errors. However, it may be that there are simply more "improved" areas in the RoUS, providing pilots with more options in case of an emergency (i.e., alternate airports, highways, roads, etc.) in which case education alone may not prove successful. Additionally, it is worth noting that "unsuitable terrain" was a classification imposed by the NTSB investigators after the fact, and the moment-to-moment judgment of how suitable terrain may be during a flight may be influenced by factors not considered fully in post hoc analyses.

Also of concern in both Alaska and the rest of the U.S. was in-flight planning/decision making. After all, decisions made during flight are often more critical than those occurring on the ground. Thus, when confronted with important decisions during flight, pilots are often under pressure to be right the first time while using limited information. Scenario-based training along these lines like that provided within the FAA-Industry Training Standards (FITS) program may improve decision-making in the cockpit, particularly if examples are drawn from the accident record.

Of the unsafe acts that aircrew commit, addressing violations may be the most difficult and complex. Recall that violations are the "willful" disregard for the rules and as such are not necessarily something that can be easily deterred or mitigated. Nevertheless, since nearly half of violations involved fatalities, behaviors like VFR flight into IMC are of great concern to the FAA and other aviation safety professionals.

Even though the percentage of accidents associated with violations did not differ markedly between Alaska and the RoUS, the specific types of violations did differ in meaningful ways. In particular, when intentional VFR flight into IMC and other adverse weather conditions were combined, an alarming 47% of the violations occurring in Alaska were accounted for (27% for the rest of the U.S.). Exactly why a larger proportion was observed in Alaska remains unknown, but one reason may be the rapid weather changes that often occur, especially around mountainous areas.

Current interventions like weather cameras in mountain passes and other locations have proved useful by providing pilots with access to real-time weather information and therefore allowing them to make informed decisions. In addition, the Medallion Foundation has provided GA pilot training using high-resolution flight simulators capable of producing simulated weather and lighting conditions and terrain depictions which are all appropriate to Alaska. With this technology, pilots are able to safely navigate through Alaska and see what flying through places such as Merrill Pass in adverse weather conditions could entail, a difficult task to successfully perform in clear conditions.

Alaska, as perhaps the FAA's largest aviation laboratory, has been the testbed for advanced avionics like those associated with the Capstone project. Enhanced weather radar, global positioning sensors, Automated Dependent Surveillance—Broadcast (ADS-B), and other cutting-edge technologies provide a more accurate picture of how the weather, terrain and traffic situation actually look from inside the cockpit. These technologies have proven useful with 14 CFR Part 135 (commuter) operations (Williams, Yost, Holland, & Tyler, 2002). However, their efficacy within GA remains to be seen.

Conclusion

In recent years, a growing concern has been directed toward GA accident rates. The FAA Administrator has set a goal of a 20% reduction in GA accidents by fiscal year 2008. If this goal is to be realized, interventions that target the underlying human causes as identified in this analysis need to be developed. Only then can any great strides in improving the GA accident rate be achieved.

References

Federal Aviation Administration. (n.d.) FAA Flight Plan 2004-2008. Retrieved August 24, 2004, from http://www.faa.gov/AboutFAA/FlightPlan.cfm.

Federal Aviation Administration. (n.d.) FAA-Industry Training Standards (FITS). Retrieved July 29, 2005, from http://www.faa.gov/education_research/training/fits/

National Transportation Safety Board. (2005). Table 10. Accidents, Fatalities, and Rates, 1985 through 2004, U.S. General Aviation. Retrieved July 12, 2005, from http://www.ntsb.gov/aviation/Table10.htm

O'Hare, D., Wiggins, M., Batt, R. & Morrison, D. (1994). Cognitive failure analysis for aircraft accident investigation. *Ergonomics, 37*(11), 1855-1869.

Shappell, S. & Wiegmann, D. (1997). Why would an experienced aviator fly a perfectly good aircraft into the ground? *International Symposium on Aviation Psychology,* The Ohio State University, 26-32.

Shappell, S. & Wiegmann, D. (2001). Beyond Reason: Defining the holes in the Swiss Cheese. *Human Factors in Aviation Safety, 1*(1), 59-86.

Shappell, S. & Wiegmann, D. (2003). *A human error analysis of general aviation controlled flight into terrain accidents occurring between 1990-1998.* (DOT/FAA/AM-03/4). Washington, DC: Federal Aviation Administration, Office of Aerospace Medicine.

Wiegmann, D.A. & Shappell, S.A. (1997). Human factors analysis of post-accident data: Applying theoretical taxonomies of human error. *The International Journal of Aviation Psychology*, *7*(1), 67-81.

Wiegmann, D.A. & Shappell, S.A. (2001a). Human error perspectives in aviation. *The International Journal of Aviation Psychology, 11,* 341-357.

Wiegmann, D.A. & Shappell, S.A. (2001b). A *human error analysis of commercial aviation accidents using the human factors analysis and classification system (HFACS).* (DOT/FAA/AM-01/3). Washington, DC: Federal Aviation Administration, Office of Aviation Medicine.

Wiegmann, D. & Shappell, S. (2003). *A human error approach to aviation accident analysis: The human factors analysis and classification system.* Ashgate Publishing Company: Aldershot, Great Britain.

Williams, K., Yost, A., Holland, J. & Tyler, R. (2002). *Assessment of advanced cockpit displays for general aviation aircraft – The Capstone Program.* (DOT/FAA/AM-02/21). Washington, DC: Federal Aviation Administration, Office of Aerospace Medicine.

PILOT ACCEPTANCE, COMPLIANCE, AND PERFORMANCE WITH AN AIRBORNE CONFLICT MANAGEMENT TOOLSET

Nathan A. Doble
Titan Corporation
Hampton, VA

Richard Barhydt
NASA Langley Research Center
Hampton, VA

Dr. Karthik Krishnamurthy
Titan Corporation
Hampton, VA

A human-in-the-loop experiment was conducted at the NASA Ames and Langley Research Centers, investigating the En Route Free Maneuvering component of a future air traffic management concept termed Distributed Air/Ground Traffic Management (DAG-TM). NASA Langley test subject pilots used the Autonomous Operations Planner (AOP) airborne toolset to detect and resolve traffic conflicts, interacting with subject pilots and air traffic controllers at NASA Ames. Experimental results are presented, focusing on conflict resolution maneuver choices, AOP resolution guidance acceptability, and performance metrics. Based on these results, suggestions are made to further improve the AOP interface and functionality.

Introduction

In today's air transportation business environment, aircraft operators are increasingly looking for means to increase flight efficiency. However, with air travel demand once again rising to levels that exacerbate delays and challenge the capacity of the National Airspace System (FAA, 2004), large efficiency improvements may be difficult to realize under current operational conditions. As a result, it has been acknowledged that a transformational, rather than evolutionary, approach to air traffic management modernization is needed (DOT, 2004).

As part of the Advanced Air Transportation Technologies project, NASA has developed such a far-term, transformational concept, called Distributed Air/Ground Traffic Management (DAG-TM) (NASA, 1999). The goals of DAG-TM are to increase efficiency and maintain safety through a redistribution of decision-making authority among airborne and ground-based elements of the air transportation system. It is a gate-to-gate concept, addressing all flight phases from dispatch to arrival.

En Route Free Maneuvering

En Route Free Maneuvering is one component of DAG-TM, addressing the en route and terminal-transition phases of flight. In an En Route Free Maneuvering environment, trained crews of equipped aircraft assume responsibility for traffic separation. Such crews would be free to modify their flight path in real time, without approval from an air traffic controller, as long as basic flow management initiatives are complied with (e.g., crossing a terminal airspace entry point at a specified time). These flights would operate under a new set of flight rules called Autonomous Flight Rules (AFR).

Except for busy terminal areas, where AFR operations would not be permitted, AFR traffic would be integrated with Instrument Flight Rules (IFR) traffic. AFR flight crews would be responsible for separation from both IFR and other AFR aircraft. Air traffic controllers would issue flow management constraints to all aircraft, and continue to provide separation among IFR aircraft, accommodating those operators who choose not to equip for AFR. By distributing separation assurance among multiple airborne and ground-based elements in this way, the National Airspace System may be able to absorb a higher increase in demand beyond what is possible with a centralized, ground-based approach.

Background

Previous Research

The work presented in this paper builds upon previous studies conducted at NASA as well as initial Free Flight research by organizations such as NLR in the Netherlands (Hoekstra et al., 2000). Past NASA experiments investigated such topics as AFR operations in confined airspace and the use of aircraft intent for decision making (Krishnamurthy et al., 2003).

The Autonomous Operations Planner

Central to AFR operations are the capabilities of airborne conflict prevention, detection, and resolution, as well as adherence to traffic flow management constraints. It is assumed that pilots cannot safely perform these functions without some form of decision support. As such, NASA Langley Research Center has developed a prototype airborne toolset called the Autonomous Operations Planner (AOP) (Barhydt & Krishnamurthy, 2004).

The prototype AOP interface is designed around a modern "glass cockpit" flight deck. It provides conflict alerts and resolution guidance via the navigation display, using state and intent data from the ownship and proximate traffic. To meet flow constraints, it also generates conflict-free paths that achieve Required Times of Arrival (RTAs) at waypoints. The AOP has been developed using a human-centered approach, with resolution guidance complementing the pilot's choice of control mode. For example, when the aircraft is being flown in a tactical mode (e.g., a selected heading or altitude) or when very near-term conflicts exist, resolution guidance is presented as a simple heading or vertical speed command. When the aircraft is flown in a strategic mode (i.e., coupled to the aircraft's flight management system (FMS)), resolution guidance is presented as an FMS route modification.

Conflicts are displayed by highlighting the intruder aircraft and indicating the region of conflict along the active flight path with a colored "dog bone." The AOP also provides information to help pilots avoid inadvertently creating new conflicts while maneuvering. These conflict prevention tools take on two forms: Maneuver Restriction Bands and Provisional Conflict Alerts. Maneuver Restriction Bands are displayed as "no fly" heading and vertical speed ranges. Using a "dashed dog bone" symbology, Provisional Conflict Alerts show regions of conflict along *proposed* flight paths (e.g., a modified but unexecuted FMS route or a selected but unengaged heading). Figure 1 shows an example of AOP symbology on a Boeing 777-style navigation display.

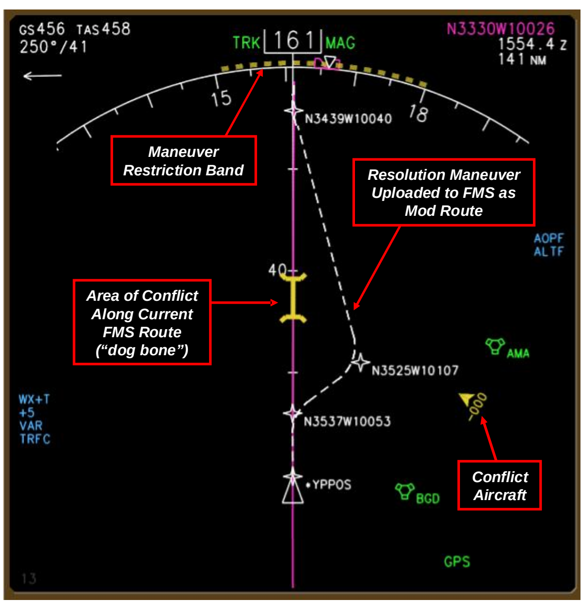


Figure 1. *AOP Interface with Strategic Resolution Experimental Approach*

In summer 2004, the NASA Ames and Langley Research Centers jointly conducted a human-in-the-loop simulation of En Route Free Maneuvering operations (Barhydt & Kopardekar, 2005). This experiment extended the previous research in several ways. A realistic, mixed AFR-IFR operating environment was simulated, including overflight aircraft as well as arrivals. The AOP was enhanced to provide vertical resolution guidance in addition to lateral guidance. In addition, interactions with ground-based air traffic controllers were studied.

This paper presents a subset of the En Route Free Maneuvering experimental results, focusing on conflict resolution maneuver choices, pilot-reported acceptability of AOP guidance, and performance metrics, including how pilot compliance with AOP affected resolution performance.

Participants

Test subjects included 12 pilots at NASA Langley as well as pilots and air traffic controllers at NASA Ames. The NASA Langley subject pilots were all Airline Transport Pilot rated with experience in Boeing glass cockpit aircraft. These pilots flew workstation-based flight simulators that emulated the displays of an AOP-equipped Boeing 777. Additional AFR and IFR background traffic was supplied with pseudo-pilot stations staffed by research personnel.

Figure 2 shows the experimental airspace. It consisted of simulated high- and low-altitude sectors of a portion of Fort Worth Center. The sectors were staffed at NASA Ames by five FAA-qualified air traffic controllers. They provided separation services between IFR aircraft and were given automated tools for conflict detection and resolution. In addition, researchers acted as pseudo-controllers in large "ghost" sectors surrounding the experimental sectors, providing limited services to flights entering and exiting the subject-controlled airspace.

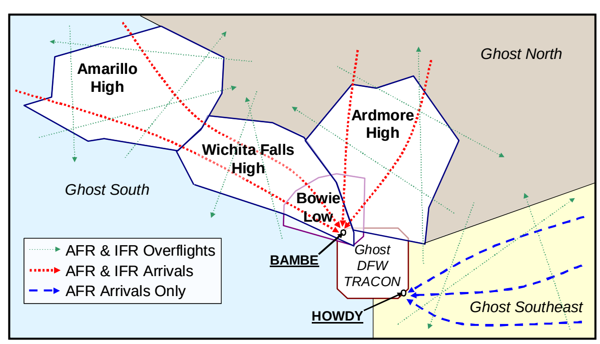


Figure 2. *Experimental Airspace*

Scenario Design

The experiment was designed in a within-subjects format, with 16 different scenarios. Four different traffic conditions were simulated, which varied the amount of traffic as well as the relative proportions of AFR and IFR overflight aircraft. Table 1 details the four traffic conditions.

Table 1. *Traffic Conditions Tested*

Condition	Avg. Traffic Density	% IFR	% AFR
C1	slightly above current Monitor Alert parameter	100%	0%
C2	equal to C1 density	75%	25%
C3	≈1.5 × C1 density	50%	50%
C4	≈2 × C1 density	35%	65%

At each of the four traffic conditions, pilots flew two overflight profiles and two arrival profiles. Except for C1 scenarios (in which all flights were IFR), subject pilots were responsible for resolving scripted and unscripted traffic conflicts. AOP alerted pilots to conflicts up to 10 minutes prior to predicted Loss of Separation (LOS). Pilots were trained to use AOP strategic resolution guidance, tactical resolution guidance, and (in the case of manual maneuvers) conflict prevention information as appropriate to the situation. They were also instructed to operate the aircraft as they would during line operations. Although hand-flying was not available, pilots were allowed to use any desired autopilot modes, including both FMS-coupled modes and tactical modes.

Results & Discussion

The NASA Langley subject pilots encountered a total of 500 traffic conflicts throughout the 12 AFR scenarios (C2, C3, and C4). For 332 of these conflicts, the subject pilot performed a resolution maneuver. The analyses presented below show results for these conflicts, without distinguishing between traffic conditions. The effects of traffic density on resolution performance are treated in a separate publication (Doble, Barhydt, & Hitt, 2005).

AOP Compliance

To examine the effects of AOP resolution maneuver compliance on resolution performance, resolution maneuvers were divided into six categories, based upon whether the maneuver was strategic or tactical and whether or not the pilot followed AOP guidance. These categories are summarized in Table 2. Two different performance metrics were then used to evaluate the maneuvers: induced conflicts and conflicts requiring multiple resolution maneuvers.

Table 2. *Resolution Compliance Categories*

Category	Description	Count
Strategic Comply	Pilot implements AOP-recommended route modification without modifications	141
Strategic Noncomply	Pilot edits waypoints before implementing AOP-recommended route modification	0
Strategic Manual	Pilot ignores or does not seek AOP resolution, and manually edits waypoints, altitudes, etc. of FMS active route	15
Tactical Comply	Pilot maneuvers in direction of AOP-recommended heading or vertical speed	118
Tactical Noncomply	Pilot maneuvers away from AOP-recommended heading or vertical speed	15
Tactical Manual	AOP tactical guidance not available, pilot implements own lateral or vertical maneuver via autopilot mode control panel	43

Induced Conflicts. The frequency of induced conflicts is a measure of the ability of pilots and AOP to account for aircraft other than the intruder when calculating a resolution maneuver. An induced conflict was defined as a new conflict arising within one minute of a previous resolution maneuver and directly caused by that maneuver. Figure 3 shows the percentage of resolutions inducing a conflict in each of the six compliance categories. Results from χ^2 tests indicate no significant differences in the frequency of induced conflicts across the three tactical categories ($\chi^2(2, N = 176) = 0.27, p > 0.05$), but a significantly higher frequency of induced conflicts for Strategic Manual maneuvers vs. Strategic Comply maneuvers ($\chi^2(1, N = 156) = 32.2, p < 0.05$).

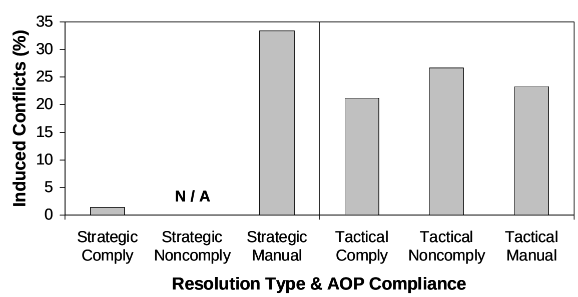


Figure 3. *Induced Conflicts vs. AOP Compliance*

The lowest induced conflict rates occurred when pilots followed AOP guidance. This highlights the advantage of decision support when resolving conflicts involving multiple proximate aircraft. It is conjectured that the relatively high rate of induced conflicts among tactical maneuvers was due primarily to two factors: the time to predicted LOS when the maneuvers were executed, and the characteristics of the AOP tactical resolution algorithm. During the experiment, tactical resolution maneuvers were generally initiated closer to predicted

LOS than strategic maneuvers. In such situations, especially in the high-density airspace simulated in this experiment, some induced conflicts may be inherently unavoidable, as the first priority is usually to resolve the most critical conflict in a timely manner. In addition, for very near-term conflicts (under 2 minutes to LOS), the AOP tactical resolution algorithm did not take other aircraft into account when calculating resolution guidance. This algorithm was chosen for its ability to successfully resolve complicated conflict situations without the need for maneuver coordination between aircraft (Eby, 1994). Ongoing research will investigate the integration of this algorithm with the AOP conflict prevention tools in order to further reduce induced conflicts.

While the significant increase in induced conflicts for Strategic Manual resolutions is cause for concern, it should be noted that three of these five induced conflicts were caused by the same pilot during the same scenario. Nevertheless, pilot training and the AOP conflict prevention symbology may warrant further attention as these subject pilots all implemented route modifications despite being shown Provisional Conflict Alerts.

Multiple Resolutions. The frequency of multiple resolutions is a measure of the ability of pilots and AOP to resolve a conflict and remain out of conflict. If a subject pilot was in conflict with the same intruder multiple times and implemented more than one resolution maneuver, this was noted as a multiple resolution conflict. Figure 4 shows the percentage of conflicts requiring multiple resolutions in each compliance category. Results from χ^2 tests indicate no significant differences in the frequency of multiple resolutions across the strategic categories ($\chi^2(1, N = 156) = 1.67$, $p > 0.05$). The differences among tactical categories were significant ($\chi^2(2, N = 176) = 6.04$, $p < 0.05$).

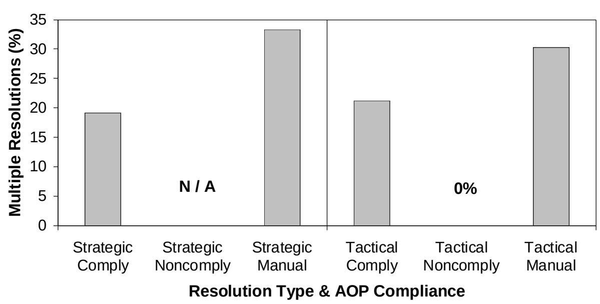


Figure 4. *Multiple Resolutions vs. AOP Compliance*

The lower multiple resolution rate for maneuvers that complied with AOP guidance (vs. manual maneuvers) shows the benefits of decision support when resolving conflicts between aircraft flying complex, four-dimensional trajectories. While the lowest multiple resolution rate occurred when pilots did not follow AOP guidance (Tactical Noncomply), this is not seen as a cause for concern. Compliance only accounted for 3% of the variance in multiple resolutions, and this category of maneuvers had a relatively small sample size. In addition, there may have been a performance tradeoff, with these maneuvers effectively avoiding the intruder aircraft at the expense of additional induced conflicts.

Choice of Maneuver Axis

To judge the relative effectiveness of lateral and vertical AOP guidance, the maneuvers categorized above as Strategic Comply and Tactical Comply were further separated into Strategic Lateral, Strategic Vertical, Tactical Lateral, and Tactical Vertical categories.

Induced Conflicts. Figure 5 shows the percentage of induced conflicts that occurred for each of the four axis categories. Results from χ^2 tests indicate no significant differences between either the strategic categories ($\chi^2(1, N = 141) = 0.46$, $p > 0.05$) or the tactical categories ($\chi^2(1, N = 118) = 0.37$, $p > 0.05$). For the reasons mentioned above, it is not surprising that strategic resolutions resulted in fewer induced conflicts than tactical resolutions, but within the strategic and tactical categories, the choice of maneuver axis appears to have had little effect.

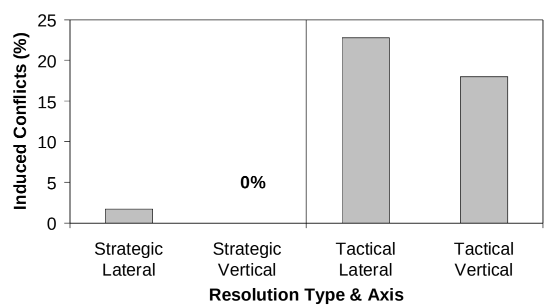


Figure 5. *Induced Conflicts vs. Maneuver Axis (When AOP Complied With)*

Multiple Resolutions. Figure 6 shows the percentage of multiple resolutions that occurred for each of the four maneuver axis categories. Results from χ^2 tests indicate no significant differences between either the two strategic categories ($\chi^2(1, N = 141) = 1.24$, $p > 0.05$) or the two tactical categories ($\chi^2(1, N = 118) = 0.02$, $p > 0.05$). This shows that lateral and vertical maneuvers were similarly effective in preventing multiple resolutions. However, the slightly higher incidence of multiple resolutions for Strategic Vertical maneuvers is worth noting. These maneuvers required pilots to adjust the autopilot altitude value in addition to uploading an FMS route modification. There were cases when the altitude value was not properly adjusted and the aircraft failed to follow the resolution maneuver. Compounding this was the difficulty of displaying vertical path changes on a horizontal situation display. Ongoing research will investigate other options for presenting vertical maneuver information, including the use of vertical situation displays.

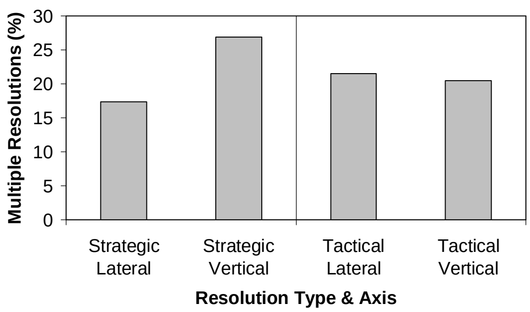


Figure 6. *Multiple Resolutions vs. Maneuver Axis (When AOP Complied With)*

Acceptability of AOP Resolution Guidance

To examine factors that affect pilot perception of the acceptability of AOP resolution guidance, subject pilots were asked after each scenario to rate, on a 1 to 7 scale, the acceptability of a) the first AOP strategic resolution in the scenario, and b) AOP tactical resolutions in general during the scenario. A series of correlations was then performed to determine if relationships existed between resolution acceptability and four other factors: conflict duration, maneuver axis (lateral or vertical), multiple resolutions, and induced conflicts. These results are presented in Table 3. Overall resolution acceptability was high for both strategic resolutions (M = 6.31, SD = 1.28) and tactical resolutions (M = 5.12, SD = 1.60).

Table 3. *Resolution Acceptability*

Attribute	Acceptability Correlation		Test
	Strategic Resolutions (N = 109)	Tactical Resolutions (N = 135)	
Conflict Duration	-0.24*	0.02	Pearson
Maneuver Axis	-0.18	0.10	Point-biserial
Multiple Resolution	-0.39*	0.14	Point-biserial
Induced Conflict	-0.02	-0.17*	Point-biserial

* = significant correlation at $p < 0.05$ level

The acceptability of AOP strategic resolution maneuvers was significantly correlated with conflict duration and multiple resolutions. The significance of conflict duration agrees with comments provided during debrief sessions, which indicated that pilots were frustrated by AOP computation delays and the options available when AOP was unable to calculate a solution. While the AOP strategic resolution algorithm (a genetic algorithm) normally converged on a solution within one second, insufficient feedback may have been provided to pilots when computation times were longer, creating the appearance that AOP had "frozen up." The significant correlation with multiple resolutions is also reasonable, as one of the primary benefits of intent-based, strategic decision support is that the necessity for multiple resolution maneuvers should be reduced by accounting for trajectory changes that would be unknown to a solely state-based system.

The acceptability of AOP tactical resolutions was only significantly correlated with whether or not the resolution induced a conflict. As mentioned above, depending on the time to predicted LOS, the AOP tactical guidance may or may not have accounted for aircraft other than the intruder. As such, there were cases when the tactical guidance disagreed with Maneuver Restriction Bands. Although this behavior was explained to subject pilots during training exercises, this is recognized as a significant human factors issue. Research is underway to modify the AOP near-term tactical resolution logic so that conflicting information is not presented to pilots.

Practice Effects

The En Route Free Maneuvering experiment lasted a total of eight days, with three days devoted to training, four days for data collection, and one day for debriefing. Each data collection day included four scenarios, with one at each traffic condition, and with the order of conditions varying across days.

To identify any learning or practice effects, conflicts were sorted by day and evaluated with the same performance metrics presented above. Figure 7 shows the frequency of induced conflicts and multiple resolutions across days. χ^2 tests indicate that no significant differences in the frequency of induced conflicts ($\chi^2(3, N = 332) = 1.37, p > 0.05$) or in the frequency of multiple resolutions ($\chi^2(3, N = 332) = 4.78, p > 0.05$) existed across the four days.

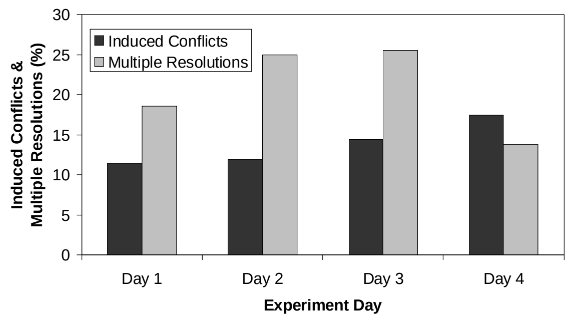


Figure 7. *Resolution Performance by Day*

While no significant practice effects were found, it is interesting to compare the performance by day with the resolution maneuvers chosen. Figure 8 shows the percentage of maneuver types chosen each day. Notionally, resolution performance appears to degrade with increases in manual and non-complying maneuvers over the first three days of the experiment, then improve on Day 4 with an increase in Strategic Comply maneuvers.

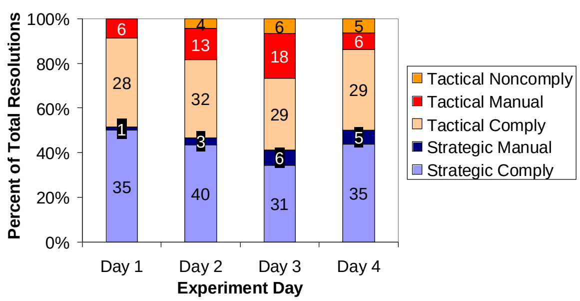


Figure 8. *Resolution Maneuver Compliance by Day*

Conclusions

Through the above analysis of conflict resolution maneuvers, several conclusions can be drawn about the performance of pilots during AOP-equipped AFR operations. First, the choice of maneuver axis (lateral or vertical) had little effect on resolution performance, indicating that resolution maneuvers can be well-executed in either axis. Second, resolution performance was shown to generally improve when pilots complied with AOP-recommended resolution maneuvers. Finally, although pilot acceptability of AOP guidance was high overall, possible ways to further increase acceptability and performance were identified. These methods include better integration of AOP near-term tactical resolution logic with conflict prevention information, improved feedback when AOP cannot converge on a strategic solution, and the potential inclusion of a vertical situation display. Along with previous findings, these results further support the feasibility of the En Route Free Maneuvering concept while highlighting areas for future research.

Acknowledgements

The authors would like to thank James Hitt at Booz Allen Hamilton for performing statistical analyses of the experimental results, and the subject pilots and controllers for their participation in the experiment.

References

Barhydt, R., & Kopardekar, P. (2005). *Joint NASA Ames/Langley Experimental Evaluation of Integrated Air/Ground Operations for En Route Free Maneuvering*. Submitted to 6th USA/Europe Air Traffic Management R&D Seminar, Baltimore, MD.

Barhydt, R., & Krishnamurthy, K. (2004). *Design of a Multi-Mode Flight Deck Decision Support System for Airborne Conflict Management.* Intl. Conference on Human-Computer Interaction in Aeronautics (HCI-AERO), Toulouse, France.

Doble, N.A., Barhydt, R., & Hitt, J.M. (2005). *Conflict Resolution Performance in an Experimental Study of En Route Free Maneuvering Operations.* Submitted to 6th USA/Europe Air Traffic Management R&D Seminar, Baltimore, MD.

DOT Joint Planning and Development Office. (2004). *Next Generation Air Transportation System Integrated Plan*. Washington, DC.

Eby, M.S. (1994). *A Self-Organizational Approach for Resolving Air Traffic Conflicts.* The Lincoln Laboratory Journal, Vol. 7, Nr. 2.

FAA. (2004). *Forecast Fact Sheet Fiscal Years 2004-2015*. Washington, DC. http://www.faa.gov /newsroom/factsheets/2004/factsheets_040325.pdf

Hoekstra, J.M., et al. (2000). *Overview of NLR Free Flight Project 1997-1999.* Contract Report NLR-CR-2000-227. National Aerospace Laboratory, Amsterdam, The Netherlands.

Krishnamurthy, K., et al. (2003). *Autonomous Aircraft Operations Using RTCA Guidelines for Airborne Conflict Management.* 22nd Digital Avionics Systems Conference, Indianapolis, IN.

NASA Advanced Air Transportation Technologies Project Office. (1999). *Concept Definition for Distributed Air/Ground Traffic Management (DAG-TM) Version 1.0.* NASA Ames, Moffett Field, CA.

SAFETY ATTITUDES IN THE AVIATION SYSTEM: INFLUENCES OF A HIGHLY REGULATED ENVIRONMENT

Teresa C. D'Oliveira
Instituto Superior de Psicologia Aplicada
Lisbon, Portugal

Although safety is considered paramount in the aviation industry, very few studies have explored the influence that such a highly regulated environment may have on safety attitudes. This paper explores how perceptions and attitudes may be influenced by context characteristics and analyses how a highly regulated context, such as the aviation industry, compares with other industries. Results suggest that the aviation industry seems to be centered on individual behaviors and attitudes towards safety; in contrast other industries highlight safety at the organizational level. Implications of these results and repercussions of national safety campaigns to promote safety at the workplace are considered.

Introduction

Workplace safety and, in particular, the analysis of occupational accidents have emphasized the importance and interrelationship of two main contributors: The technical component which involves physical working conditions, machinery, equipment and work instruments and the Human component comprising job incumbents, teams, supervisors and top managers (e.g., Oliver, Cheyne, Tomás & Cox, 2002; Sarkus, 2001). The development of a positive safety culture and constructive attitudes towards safety are considered as an important and effective strategy to promote and maintain a safe workplace. In many instances, attitude surveying is recommended as a quick and helpful way of conducting a safety diagnostic.

The literature on safety attitudes presents a variety of dimensions and a plethora of instruments (e.g., Cox and Cox, 1991; Díaz & Cabrera, 1997; Glendon, Staton & Harrison, 1994; Zohar, 1980). A renaming and grouping exercise on the existent measures is considered necessary (Guldenmund, 2000; Sorensen, 2002) with possible identification of core dimensions and clear explanations on the issue of dimensionality. These efforts led to the development of a measure to evaluate attitudes towards safety that can be used in various contexts (D'Oliveira, 2004). A methodology similar to the one adopted by Williamson, Feyer, Cairns and Biancotti (1997) was used and a measure considering eight scales was put together. Safety areas considered were: Organizational objectives, organizational practices and safety, information on safety issues, management and supervisors' attitudes, personal attitudes to safety, risk perceptions and relationships with co-workers.

Safety is paramount in the aviation system and efforts have considered both the technical component (e.g., by fostering safer machinery) and human interventions (e.g., through improved training like CRM). The industry investments in standards and practices led to an outstanding safety record (ICAO, 2004).

Context characteristics such as the activities performed, the hazards involved and the degree of regulation imposed by the industry may play an important role when discussing safety attitudes and safety culture. These characteristics have yet to be considered in the literature on safety culture/climate. Very few studies have considered safety attitudes in different industries (e.g., Diaz & Cabrera, 1997). This paper addresses these issues and explores how perceptions and attitudes may be influenced by context characteristics and analyses how a highly regulated context, such as the aviation industry, compares with other industries.

Method

Participants

A total of 346 participants, 60.4% men and 396 females, from various industries (aviation, health, car industry, metal industry, etc.) were invited to participate in this study. Table 1 presents sample's main characteristics.

Table 1. *Participants' main characteristics*

Age	M = 36.71, SD=10.09
Qualifications	M = 9.9 years
Contract	Full time permanent = 84.3 %
Position	Supervisor = 19.1%
Industry	Aviation = 25.4%; Non Aviation = 74.6% Pilots, Cabin crew, Maintenance

Instrument

A measure was developed using a methodology similar to the one adopted by Williamson, Feyer, Cairns and Biancotti (1997) was used in this study. Specifically, a review of the literature was conducted in order to identify potential measures of attitudes towards workplace safety. All potential measures were then considered as a full set and items were assembled according to their content. This procedure led to the identification of seven dimensions: organizational objectives, organizational practices and safety, information on safety issues, management and supervisors' attitudes, you and safety issues, personal appreciation of risk and relationships with coworkers. A detailed definition of each dimension (Table 2) was then produced and eight items were selected to represent each safety attitude dimension. The final measure was composed of 56 items, each item being responded in a 5 point rating scale.

Table 2. *Safety attitude dimensions (Cronbach's values for each dimension).*

Sub-Scale Definition
A – Organizational Objectives This dimension considers how the Organization values safety issues. The potential conflict between safety and productivity, the Organization openness to discuss issues related to safety and proposals by the employees are some of the issues considered in the literature (α = .725).
B- Organizational Practices & Safety This dimension addresses how organizational practices such as training, performance evaluation, promotion, accident/incident investigation may be related with safety (α = .850).
C - Information on Safety Issues This dimension tries to evaluate how the Organization stimulates the diffusion of information related with safety by creating safety awards, safety bonus, how workers might present suggestions or report their safety concerns, etc (α = .720).
D- Management & Supervisors Attitudes In this dimension, supervisors and top managers' behavior is considered by assessing workers perceptions of their technical knowledge on safety issues, proactive or reactive safety attitude and their support to workers safety concerns (α = .806).
E – Yourself & Safety This dimension considers the knowledge and satisfaction of workers in relation to safety and their awareness of the consequences of their behavior to safety in general (α = .776).
F – Risk Perceptions In this dimension workers' perceptions of the risks involved in their activities are considered along with their estimative of how probable it is to be involved in an accident (α = .717).
G – Relationships with coworkers This dimension considers workers perceptions of their colleagues' knowledge and behaviors related to safety. It also includes the perception of being part of a group and how this characteristic influences personal behavior (α = .808).

Procedure

A general instruction was given to every participant as to how they should fill in the questionnaire: volunteers should give a description of their own company regarding safety issues. The objective of the study was to gather information that could help companies to improve their safety policies and results.

Results

A total of seven MANOVAS were conduct in order to explore potential differences between aviation and non aviation participants. Table 3 summarizes main results obtained in these analyses.

Table 3. *Differences between aviation and non-aviation participants in each subscale*

Sub-Scale	Results
Organizational Objectives	Pillai's Trace= .163, F= 7,990; p<.000 Non aviation has higher means
Organizational Practices & Safety	Pillai's Trace= .108, F= 4,958; p<.000
Information on Safety Issues	Pillai's Trace= .135, F= 6,374; p<.000 Non aviation has higher means (ns differences)
Management & Supervisors Attitudes	Pillai's Trace= .141, F= 6,859; p<.000
Yourself & Safety	Pillai's Trace= .266, F= 15,149; p<.000 Aviation has higher means
Risk Perceptions	Pillai's Trace= .105, F= 4,751; p<.000 Non aviation has higher means
Relationships with coworkers	Pillai's Trace= .189, F=9,675; p<.000

Discussion

Results obtained suggest differences between aviation and non-aviation participants in every dimension. In what concerns organizational objectives, information on safety issues and risk perceptions, non-aviation systematically has higher means.

Non-aviation participants compose a positive depiction of their companies: safety goals are clearly stated, safety procedures work well and are followed, there seems to be more information available on safety issues but it is recognized that sometimes there is a conflict between productivity and safety, something mentioned in the literature.

In what concerns aviation participants, they seem to have a personal relation with safety issues that appears to be more positive (receive safety information, understand safety rules, know training needed) and there is a proactive attitude towards safety (recognize that their personal intervention may avoid potential hazards), and attitudes and behaviors associated with an appreciation of risks involved in their jobs.

Organizational practices and procedures towards safety, management and supervisor's behaviors and attitudes and relationship with colleagues although presenting mixed results provide support to the differences previously identified. In the aviation context, safety is part of performance appraisal, supervisors are aware of what safety training each worker has, and participants report reliable safety behaviors in their colleagues. Non-aviation participants report that their work procedures are accurate and a reflection of what they actually do in their jobs, characteristics probably associated with a lesser degree of complexity in their jobs.

All in all, results suggest the presence of two different safety systems. Aviation safety systems seem to have at their centre individual safety qualifications: a greater risk in the activities performed is associated with the requirements for specific and formal safety training. Such qualifications are quite relevant in this context; not only are they included in the performance appraisal but also management and supervisors are aware of each worker qualifications. In the aviation context, if you do not have the necessary safety training, you will not be able to work.

In contrast, non-aviation industries seem to centre on the company safety records as a whole: company goals are emphasized, general information on safety is available, supervisors encourage involvement in safety issues and are perceived to know safety inspections' results. This analysis is further supported by non-aviation better results in "we are recognized and rewarded for working together".

In this sense it would be appropriate to say that aviation safety systems are individualistic by nature and non-aviation safety systems are much more collectivistic. Such perspectives can also be associated with an "organizational locus of control or accountability".

Results from non-aviation organizations may be related with recent government investments in workplace safety. Portugal has one of the worst work accident rates in the European Community. Support for safety training, safety programs, safety prizes, safety inspections and media campaigns have been created to address this problem. The problem is depicted as a national problem (national statistics may involve anyone) or an organizational problem (fines for companies that do not follow safety recommendations) and an issue that needs every person's contribution. Such perspective helps to depart from an individualistic approach of work accidents or the bad apple theory (Dekker, 2002) that hinders organizational safety learning. Advantages of this viewpoint should be considered by aviation safety systems as it may complement the existing perspective.

References

Cox, S.J. & Cox, T.R. (1991). The structure of employee attitudes to safety: An European example. *Work & Stress*, 5, 93-106.

D'Oliveira T. (2004). *Atitudes face à segurança: um exercício de reagrupamento e redenominação da literatura* ("Attitudes towards safety: a regrouping and renaming exercise on the literature). Paper presented at the 2º Congresso Hispano-Português de Psicologia. Universidade de Lisboa, Portugal (22-25 Setembro 2004).

Dekker, S. (2002). *The Field Guide to Human Error Investigations*. Aldershot, UK: Ashgate.

Díaz R. I. & Cabrera, D. D. (1997). Safety climate and attitude as evaluation measures of organizational safety *Accident Analysis & Prevention*, 29(5), 643-650.

Glendon, A. I., Staton, N. A. & Harrison, D. (1994). Factor analyzing a performance shaping concepts questionnaire. In Robertson, S. A. (Ed.), *Contemporary Ergonomics 1994: Ergonomics For All* (pp. 340-345). London, UK: Taylor and Francis

Guldenmund, F. W. (2000). The nature of safety culture: a review of theory and research. *Safety Science*, 34, 215-257

ICAO (2004). *Cross-cultural factors in aviation safety*. Human Factors Digest, nº 16, Montreal, Canada: ICAO

Oliver, A., Cheyne, A., Tomás, J. M. & Cox, S. (2002). The effects of organizational and individual factors on occupational accidents. *Journal of Occupational and Organizational Psychology*, 75, 473-488.

Sarkus, D. J. (2001). Safety and psychology: where do we go from here?, *Professional Safety*, January, 18-25.

Sorensen, J. N. (2002). Safety culture: a survey of the state-of-the-art. *Reliability Engineering and Systems Safety*, 76, 189-204

THE EFFECTS OF A SCENARIO BASED GPS TRAINING PROGRAM ON PILOT PROFICIENCY IN THE GENERAL AVIATION PILOT

Wayne A. Dornan
John Bertrand
Paul Craig

Middle Tennessee State University, Department of Aerospace, 1500 Greenland Drive, Box 67, Business and Aerospace Bldg., Murfreesboro, Tennessee 37132

Currently, General Aviation (GA) pilots working toward their instrument rating (IFR) in aircraft equipped with Global Positioning System (GPS) technology receive little, if any, formal flight instruction on GPS technology. Therefore, the hypothesis examined in this study was that instrument rated pilots already certificated to fly IFR / GPS have insufficient knowledge of the GPS technology to use it effectively. Our goal was to develop a single pilot crew, scenario-based training program to increase the knowledge and safety of pilots using this technology by focusing on GPS mode awareness, situational assessment, risk and time management, and situational awareness.. The study included thirty-four pilots who had completed their instrument rating in a GPS-equipped aircraft within the last 12 months. All participants were given Pre-experiment GPS screening tests to assess overall GPS knowledge and, more specifically, knowledge regarding the Garmin 430. Each participant underwent simulator familiarization sessions in a Frasca 142 flight simulator equipped with a panel mounted IFR approved GPS. After the familiarization sessions, participants were randomly assigned to one of two groups: 1) Experimental or 2) Control. All groups then flew IFR-generated flight scenarios designed to assess their aircraft system monitoring skills (situational assessment), GPS mode awareness, situational awareness, and understanding of the Garmin 430 IFR programming. Each scenario lasted approximately 60 minutes. Following the first session, the experimental group received training sessions concentrating on mode awareness, situational awareness, time management, and situational assessment using computer based training (CBT) with a Garmin 430 simulation software program. The control group received sessions that covered basic IFR flying skills. Following the training sessions both groups flew another scenario in the Frasca 142 simulator. Each subject was video-taped to assess eye fixation in three areas of interest: 1) out the window, 2) aircraft instruments, and 3), GPS display. The results of this study revealed that a GPS scenario-based training program significantly reduced omission errors and incorrect or inappropriate use of the GPS when compared to controls. In addition, a significant effect of training versus eye gaze was observed. Pilots in the experimental group spent significantly less time gazing at the GPS following the training sessions and more time gazing at the instruments compared to controls.

Introduction

Over the last decade, there has been profound development in regard to automated flight deck technology, undoubtedly leading to increased safety in commercial aviation (Parasuraman and Riley, 1997). Until very recently, however, issues with automated flight decks were only relevant to the commercial flying industry involving multiple flight crews (Endsley and Kabor, 1999; Funk and Lyall, 1997; Sarter and Woods, 1995). This is no longer the case with the advent of automated cockpits that have recently proliferated in the General Aviation (GA) community.

One critical component of any automated cockpit is its flight navigation system (Wiener, 1988). One of the most popular in the GA community is the Global Positioning System (GPS). In fact, it is estimated that as many as two thirds of GA pilots use some form of GPS technology to navigate (St. George, 2000). Currently, in the GA community, however, there is no accepted training program for aircraft equipped with GPS technology. Indeed, this has led to a kind of "self-instruction" where GA pilots either teach themselves to use their GPS or obtain informal instruction from other GPS users. While in some cases this has resulted in only minor problems, in other cases, the results have been more tragic (O'Hare and St. George, 1994). Consequently, one key issue with the establishment of GPS technology in the GA aircraft is how to train pilots/students to take advantage of the increased safety opportunities available with the new technology. This is in striking contrast to the commercial airline industry where there is a plethora of scenario-based training programs involving specific events known to be problematic for multiple crew in an automated flight deck (Funk, Lyall, and Niemczyk, 1997).

Based on the above, it should come as no surprise that the general consensus emerging among the GA community, as well as the FAA, is that a thorough training program is needed to educate pilots on the use of GPS technology. Indeed, in a recent study on GPS usability, Adam at el. 2004, recommend that a specific GPS training program be compared to a control group not receiving any formal GPS training. If successful, the training program could be submitted to the FAA for incorporation in flight schools (Adam et al. 2004). Currently, however, the authors are unaware of any empirical data that exist to support the notion that a specific GPS training program will increase pilot proficiency in the use of GPS technology. Moreover, what constitutes a viable training program is also unknown.

Methods

This study was comprised of pilots who had completed their instrument rating in a GPS-equipped aircraft within the last 12 months at MTSU. All participants were given screening tests to evaluate their overall GPS knowledge as well as their specific knowledge of the Garmin 430. In addition, all participants were given a flight questionnaire regarding demographics, flight experience, comfort flying alone in the IFR environment, flying alone IMC, and personal IFR minimums. Before the beginning of the experiment, each participant was given a familiarization session in a Frasca 142 flight simulator equipped with a panel mounted IFR approved GPS located in the MTSU Department of Aerospace. During these sessions pilots were instructed to fly an instrument approach into Nashville International airport without the GPS. After the familiarization sessions, participants were randomly assigned one of two groups: 1) Experimental or 2) Control. All groups then flew two IFR-generated flight scenarios that were designed to assess their aircraft system monitoring skills (situational assessment), GPS mode awareness, situational awareness, and understanding of the Garmin 430 IFR programming. Each scenario lasted approximately 60 minutes. Following the first session, the experimental group received training sessions concentrating on mode awareness, situational awareness, time management, and situational assessment using a CBT with a Garmin 430 simulation software program currently available in the Aerospace Department. In contrast, the control group training sessions covered basic IFR flying skills. Following the training sessions both groups flew another scenario in the Frasca 142 simulator. During the scenario-based flying sessions, incorrect or correct GPS mode usage was recorded. A score of “0” was assigned to activities that required GPS programming following an ATC instruction, but were omitted by the pilot. A score of “1” was assigned to activities that required GPS programming following an ATC instruction, but where the GPS was used inappropriately. A score of “2” was assigned to activities where appropriate GPS use occurred, but the pilot failed to comply with an ATC instruction. A score of “3” was assigned to activities where accurate use of the GPS along will full ATC compliance was observed. For example, in each scenario, pilots were instructed to descend and cross a particular “fix” at a specific altitude. A score of “3” was recorded if the pilot used the “VNAV/VSR” in the GPS (an appropriate GPS mode). If the pilot only began to descend, a score of “0” was recorded. Participants were video taped in order to monitor the overall scanning patterns in the cockpit. Using a template developed by Diez et al. (2001) data analysis of eye fixation was based on dwell time in three areas of interest (AOI): 1) out the window, 2) aircraft instruments, and 3), GPS display. Following the completion of the study, all parametric data were analyzed using an analysis of variance (ANOVA) mixed design. Any significant main effects were assessed by Post Hoc analysis using the Scheffe’s test. Non-parametric data was analyzed using the Kruskal-Wallis test.

Results

As can be seen in Table 1, a multivariate comparison of group means of total instrument time, total time, and total actual time, revealed non-significant differences between the experimental group compared to controls ($p > 0.05$).

GROUP	AGE	Total Time	Total Instrument	Total Actual
CTL (n=17)	20	181	43	4.2
EXPL (n=19)	2	220	42	2.3

Table 1. *Mean age, total flight time accumulated, total instrument time, and total actual flight time in the experimental group compared to controls.*

Table 2 illustrates GPS competency scores in the two groups following a specific GPS CBT program (experimental group), compared to controls (traditional IFR training). A 2X2 factorial design revealed a significant group by treatment interaction $F = 29.6\ (1,35)$, $p < 0.01$. Following the CBT scenario-based training, the experimental group made significantly fewer errors compared to controls (see Table 2). In addition, as can be seen from Table 3, a 2X2X3 ANOVA revealed a significant 3 way Group by Session by AOI interaction on eye gaze in the experimental group compared to the controls; $F = 32.89\ (2,198)$ $p < 0.01$. Participants who had the GPS scenario-based training seminar (experimental group) spent significantly less time gazing at the GPS unit and more time on the Flight/Engine instruments compared to controls.

GROUP	BEFORE TREATMENT	AFTER TREATMENT
EXPERIMENTAL	7.8 +. 76	29.6 + .9 ***
CONTROLS	4.1 + .65	11.4 + 1.3

Table 2. *Means + standard error of the mean of the mean. GPS errors in the experimental group that received GPS scenario-based training compared to controls that received IFR training on two simulator flight scenarios. *** = p < 0.01 A higher score corresponds to more errors .*

GROUP	WINDOW	INSTR	GPS
EXPERIMENTAL	4.23 + .1	22.3 + .3	43 + .2 ***
CONTROL	6.68 + .1	48 + .1***	5 + 1.2

Table 3. *Mean eye gaze + standard deviation of the mean in the experimental group that received GPS scenario-based training compared to controls that received IFR training. *** = p < 0.01*

Table 4 illustrates the results of several questions that were asked regarding IFR experience, as well as personal IFR minimums. As can be seen from Table 1, the mean total actual time is surprisingly low. Another surprising result can be seen in Table 4. When asked about personal IFR minimums, an overwhelming majority (76%), reported that "they had never really thought about it". A Spearman's correlation coefficient between total IFR hours and personal minimums revealed no significant relationship.

TOTAL IFR TIME	MEAN VISIBILITY	MEAN CEILING	PERCENT NEVER THOUGHT ABOUT IT
>151 HRS	1.86 + .2	1200 + 299	68 %
101-150	1.6 + .5	1329 + 273	71 %
35-100 HRS	1.3 +	1000 + 300	75 %

Table 4. *Reported Mean visibility and Cloud minimum reported based on IFR flight experience. In addition to minimums, the percentage of participants that had "never thought about personal minimums" is reported.*

Discussion

The results of this study revealed that prior to a GPS training program, pilots who were perfectly legal to fly an IFR GPS panel mounted aircraft knew very little about the procedures involved. This was reflected in a significant amount of inappropriate programming, omission errors where the GPS was not used following an ATC clearance, poor time management, and lack of mode and situational awareness. This lack of GPS awareness resulted in a significant amount of time spent pre-occupied with the GPS which resulted in a lack of situational awareness (many were completely disoriented and, as a result, often dangerously off course), as well as situational assessment (most spent a significant amount of time focusing on the GPS and considerably less time monitoring engine/flight instrument panel). For example, in many cases the over-focusing on the GPS display resulted in altitude busts or overshooting a heading following an ATC assigned vector.

The results of this study reveal that a maneuver based approach characteristic of traditional IFR training is insufficient given the dramatic changes in technology that now typify GA aircraft. Moreover, the traditional IFR training curriculum with a focus on rote learning, is in our opinion, also antiquated and must be changed. For example, now where in the current Part 141 syllabus is GPS mode awareness, or proper time management skills emphasized. Typically the focus is on learning how to fly a GPS approach using multiple approaches.

In this study, the experimental group which received CBT seminars focusing on scenario-based training had significantly fewer errors compared to controls that utilized the traditional focus on IFR maneuvers. In our opinion, all curricula which utilize aircraft with GPS technology should incorporate at least four major components. First, ground school should focus on GPS technology and specific GPS knowledge regarding the equipment available in the aircraft, followed by specific tests to assess the students' knowledge. Second, the flight training should incorporate realistic GPS scenario-based training utilizing CBT. Indeed, CBT has the advantage of enabling the instructor and student to focus on such critical tasks as time management, proper mode awareness, and situational awareness. Third, a minimum of five hours should be required for simulator training using realistic scenarios immediately following CBT. Here, the focus would be on incorporating system management, mode awareness, and situational assessment while actually flying the aircraft. Lastly, a specific checklist should be developed that emphasizes the technology that is on board the aircraft. In our study, we developed a specific checklist that emphasized GPS mode awareness. For example, the last item on the before takeoff checklist was: "GPS / OBS………AS REQUIRED". This was designed to prompt the pilot to consider what was the appropriate GPS mode for takeoff.

In conclusion, the results of this study reveal that a GPS scenario CBT based training program significantly reduces omission errors and incorrect or inappropriate use of the GPS when compared to controls. The added benefit of this training program is pilots then spent significantly less time gazing at the GPS panel and more and more time gazing at the instruments.

References

Adam, H., Deaton, J.E., Hansrole, R.W., Shaikh, M.A. (2004). An Analysis of GPS Usability and Human Performance in Aviation. International Journal of Applied Aviation Studies. 4, 107-131

Amalberti, R.R. (1998). Automation in Aviation: A human factors perspective. In D.J. Garland, J.A.Wise, and V.D. Hopkin, (eds), Handbook of Aviation Human Factors (pp.173-192). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Diez, M., Boehm-Davis, D.A. Holt, R.W. (2001). Tracking Pilot interactions with flight management systems through eye movements. Handbook of Aviation Human Factors (pp.121-140). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Endsley, M.R. and Kabor D.B. (1999). The out-of-the-loop performance problem and level of control in automation. Human Factors, 37, 381-394.

Funk, K.H. and Lyall, E.A. and Niemczyk, M.C. (1997). Flight Deck automation Problems: Perceptions and Reality. In M. Mouloua and Koonce, J.M. (Eds). Human-automation interaction: Research and Practice. Mahwah, New Jersey: Lawerence Erlbaum Associates.

Huettig, G., Anders, G. and Tautz, A. (1999) Mode Awareness in the modern glass cockpit attention allocation to mode information. In R. Jensen (Ed), Proceedings of the 1999 Ohio State Aviation Psychology Conference. Dayton, Ohio. Ohio State University.

Kintsch, D.E., and Greeno, J. (1985). Understanding and solving word arithmetic problems. Psyhological Review, 92, 109-129.

National Transportation Safety Board. (1999) Aircraft Accident Report, Controlled flight into terrain, Korean Air, Flight 801, Boeing 747-300, HL7468, Nimitz Hill, Guam, August 6,1997. (Report Number: AAR-99-02). Washington, D.C. National Transportation Safety Board.

O'Hare, D. and St.George, R. (1994). GPS- (Pre) Cautionary Tales. Airways. 6, 12-15.

Parasuraman, R., and Riley, V. (1997) Humans and Automation: Use, misuse, disuse and abuse. Human Factors, 39, 230-253.

Riley, V. (1996) Operator reliance on automation: Theory and Data. In R.

Parasuraman and M. Mouloua (Eds). Automation and human performance: Theory and Applications (pp.19-35). Mahaw, New Jersey: Lawrence Erlbaum Associates.

Sarter, N.B. and Woods, D.D. (1995). How in the World did we ever get into that mode? Mode error and awareness in supervisory control. Human Factors, 37, 5-19.

SYNERGY OF VIRTUAL VISUAL AND AUDITORY DISPLAYS FOR UAV GROUND CONTROL STATIONS

Susan R. Dowell and **R. Jay Shively**
Army/NASA Rotorcraft Division
Aeroflightdynamics Directorate (AMRDEC)
US Army Research, Development & Engineering Command
Moffett Field, CA

Unmanned aerial vehicle (UAV) operators must remotely manipulate payload sensors, while maintaining situational awareness from a displaced ground control station (GCS). Potential use of helmet-mounted displays (HMD) in piloting UAVs and controlling payload sensors has been previously investigated (de Vries & Padmos, 1997; Draper, Ruff, & LaFleur, 2001; Morphew, Shively, & Casey, 2004). Stated benefits of HMD use for targeting tasks included immersion in the search environment and possible reduction of tactical footprint. In the current study, it was hypothesized that the pairing of 3-D audio alerts with the HMD would result in more robust performance differences between HMD and CRT conditions. For this experiment, eight subjects conducted routine area searches, periodically responding to audio threat alerts. Audio alerts were given in mono, stereo, and 3-D spatialized presentation. Targeting performance differences were assessed in a baseline CRT and joystick configuration versus HMD for all audio conditions. Findings revealed more precise target acquisition performance when payload operators used the CRT/joystick configuration than the HMD. Furthermore, time on target was reduced when visual searches were aided with stereo and 3-D directional audio cues. Lastly, participants missed the fewest targets and reported lowest workload levels, when receiving 3-D audio cues. Present findings replicated reported sickness associated with HMD use. A synergistic effect of 3-D audio and HMD showed a mitigation of operator workload previously reported with the HMD. Further consideration of 3-D audio alerting for UAV operators should be investigated for benefits in target acquisition, reduced operator workload, and increased situation awareness.

Introduction

Unmanned aerial vehicle (UAV) ground control stations present the unique environment of displacing the operator from the vehicle flying. This displacement removes typical cues used by pilots (e.g., proprioceptive, visual, vestibular) to aviate effectively and maintain situational awareness. Current unmanned aerial vehicle ground control stations are characterized by traditional workstation layouts: Two multi-function displays per station, a keyboard, and joystick (e.g., Shadow, Predator). Synergy of 3-D audio alerting with an HMD may lead to higher target acquisition performance, increased situational awareness, and lower operator workload than the current interface.

Potentially, helmet-mounted displays (HMDs) offer a reduced system footprint and the benefit of an immersive search environment for the mission payload operator (MPO). Thus far, empirical data has revealed only limited success with HMDs in the control of UAV payload sensors (Draper, Ruff, Fontejon, & Napier, 2002; Morphew, Shively, & Casey, 2004). Noted caveats for HMD use have been associated with visual lag (Rash & McLean, 1999), head-coupled sensor manipulation (de Vries & Padmos, 1998), and potential sickness side effects (DiZio & Lackner, 1997). Consequently, improvements are necessary to obviate the reported costs associated with HMD use and possibly contribute with a reduced tactical footprint.

By way of improvement, guided visual searches eliciting slower head movement may mitigate previously reported operator discomfort. As such, the presentation of aural target information that is spatially localized, or 3-D audio cues, may facilitate more efficient visual searches for air and ground targets. This has been shown in cockpit applications to enhance the acquisition of air traffic, targets and incoming threats (Begault & Pittman, 1994). Benefits of 3-D audio in presenting target location information and threat avoidance have also been reported in simulated military applications where ambient noise in the cockpit competes with audio signals (Ericson, 2004). In applying these findings to UAV ground control stations, 3-D audio technology may similarly enhance operator performance.

The current experiment combined 3-D audio alerting with an HMD to assess the potential benefits to UAV operators on nominal search missions. For the purpose of comparison, mono and stereo cueing were also employed to assess the impact of spatialized target location information. Current ground control station configuration featuring CRT and joystick was used as a baseline for display presentation. Audio alerts were presented in both HMD and CRT display environments. Findings were expected to reveal a significant interaction between display type and audio alert condition, such that using an HMD with 3-D audio alerting yielded best operator performance. Additionally, operator workload was anticipated to decrease relative to expedited searches, directed by 3-D audio cueing.

Method

Participants

Eight right-handed, male participants between 18- 30 years old (*M* = 24 yrs.) with normal or corrected-to-normal vision and full ability to perceive color were tested in this experiment. All participants reported no hearing impairment. Monetary compensation was given for participation in the study.

Simulation Equipment

CRT and Flybox. Participants were tested in a UAV simulator based on the US Army's Tactical UAV (Shadow) Ground Control Station. The simulated sensor payload view was displayed on either a CRT or HMD, depending on experimental display condition. When the sensor view was displayed on the CRT, a 21" Silicon Graphics color monitor was located 65 cm from the participant's vantage point. Display resolution was 1024 x 768 pixels. UAV sensor heading and pitch were driven by the participant's manipulation of a spring-centered joystick on a BG Systems Flybox. In an attempt to simulate the U.S. Army's TUAV sensor payload, joystick manipulation enabled 360 deg pan capability with +45 to -115 deg pitch limitations (U.S. Department of the Army, 2001). Sensor slew rate operated at a constant 60 deg/sec.

HMD and headtracker. Alternately, the sensor view was displayed on a Kaiser ProView™ XL50 head mounted display, featuring a 30 deg vertical x 40 deg horizontal FOV with 100% binocular overlap (Figure 1). Display resolution was 1024 x 768 pixels. A Polhemus Fastrak electromagnetic head tracker transmitter was mounted on the HMD and used to track subjects' head movement. In this manner, subjects' head movement was coupled to sensor movement (i.e., turning the head left moves simulated sensor view left). Head movement manipulated sensor movement in x-y and pitch axes. As in a previous study (Morphew, Shively, & Casey, 2004), the sensor view contained an artificial 45 deg downlook bias. The built-in bias afforded an optimal 45 deg sensor downlook angle when the subject's chin was parallel to the ground and his eyes on the horizon, without necessitating a fatiguing sustained downward head tilt. Sensor slew rate matched physical limitations of the TUAV sensor (max. 60 deg/sec.). Consequently, head swivel movement actuated a sensor slew movement of no greater than 60 deg/sec. When head swivel movement exceeded 60 deg/sec., a programmed limiter was engaged, allowing for no greater than a 60deg/sec. pan capability. Graphics presentation and data collection were updated at 30 Hz for all display conditions.

Figure 1. *Kaiser ProView™ XL50 HMD with headtracker and flybox.*

Audio equipment. For all experimental trials, AuSim software generated audio alerts delivered through a Sennheiser HD570 headset. When delivering 3-dimensional audio alerts, spatialized sound was referenced to the participant's head position, which was calculated by the headtracker. In this manner, alerts were generated that sounded as if they originated from a point in space. Sampling of head position associated with localized alerts was updated at 30 Hz.

Simulation

Environment. The experimental scenario simulated an area reconnaissance conducted by a mission payload operator. The virtual scene displayed on either the CRT or the HMD was analogous to the sensor video feed from a notional tactical UAV. Medium-resolution, charcoal gray roads overlaid mottled brown terrain with some instances of green shrubbery and trees alongside the roads. Portions of the flight route were located in more populated areas of the database, which included buildings and other cultural features. Desert-camouflaged tanks (targets) and green-camouflaged tanks (non-targets) were positioned throughout the simulated environment (Figure 2). Placement of vehicles throughout the simulated terrain varied according to which of the eight nominally similar flight routes were flown.

UAV control and flight path were pre-programmed and operated in playback mode throughout the simulated missions. All mission scenarios were flown at 70 KIAS and an altitude of 5000 ft. AGL.

Figure 2. *Target (left) and non-target (right).*

Audio cues. Assuming complete accuracy of an automatic target recognition system, audio cues alerted subjects to the presence of a target. One of three types of audio cues was presented, depending on experimental condition. Audio cue type was characterized as Mono (non-directional), Stereo (left-right localization), or 3-D localized format. In the Mono audio cue condition, alerts were given in both ears of the headset. The alert consisted of a female voice repeating, "Target. Target." for a duration of 10 seconds, or until the target was identified. The alert ceased once the target was identified or if undetected, at the end of the 10 sec. window.

In the Stereo audio cue condition, alerting was presented in the left or right ear according to the location of the target relative to the current sensor heading. When utilizing the HMD for sensor control, stereo audio alerting was also relative to the head direction, as sensor position was coupled to head position. The content and duration of the alert was identical to the audio cue used for the Mono audio cue condition.

For the 3-D localized cue condition, alerts were given in spatialized presentation to the left or right ear and continuously updated with sensor/head position. Due to the nature of spatialized sound, audio cues appeared to originate in 3-D space, co-located with the target position. Accordingly, alerts could shift from left to right ear as updated to sensor position and referent to relative target position. Content and duration of the 3-D audio cues were identical to those detailed for all other cue conditions.

Search Task

A routine area search was conducted in each mission scenario. The mission instructions dictated that all vehicles found were classified as targets or non-targets. A button on the flybox was used to mark non-targets, while a trigger on the joystick marked targets. Marking of non-targets during periods without audio alerts served as a secondary task to prevent boredom and preserve vigilance by maintaining a level of work. Participants were instructed to immediately respond to any audio alert by moving the sensor in the direction of the audio cue, until the target was in sight. In experimental trials with mono audio alerts, the subject did not have directional information and therefore had an unguided search for the target. In all other audio cue conditions, the target search was guided. Once the target was detected, participants centered the target within superimposed crosshair symbology and depressed a trigger on the joystick. Subjects were instructed that targeting accuracy and speed were equally important. Targets were only visible during the time of the alerting. Otherwise, targets disappeared upon trigger depression or at the conclusion of the audio alert. After acquiring the related target, subjects returned to the secondary task of marking non-targets. A total of 12 targets were presented in every mission. Mission duration was approximately 12 minutes.

Experimental Design

A within-subjects design with repeated measures was conducted. The independent variables investigated were Display type (CRT or HMD) and Audio type (Mono, Stereo, or 3-D). Subjects participated in two sessions each for a total of 6 hours per subject. Separate sessions were necessary to isolate effects of display condition (HMD, CRT). The sequence of display testing was counterbalanced. Audio type alerts were blocked and randomized within the display condition. Two replications of each Display (2) x Audio type (3) mission were completed, for a total of 12 missions or 6 per session (Figure 3).

Subject ID	Day	Display	Audio			# of Trials
1	1	CRT	3-D x 2	Mono x 2	Stereo x 2	6
	2	HMD	Mono x 2	3-D x 2	Stereo x 2	6
					Grand Total =	**12**

Figure 3. *Experimental design with Subject 1 as exemplar.*

Data Collection

Targeting acquisition: Speed and accuracy. Objective performance measures included speed and accuracy of target acquisition. Speed of acquisition was calculated from the onset of the audio alert to the subject's trigger depression. Speed was measured to the nearest hundredth of a second. Accuracy of target acquisition was measured in pixels from the center of the superimposed crosshair symbology to the centroid of the tank. Targeting error was calculated in real-time data collection. All missed targets were recorded.

Workload ratings. The NASA-TLX subjective ratings scale (see Hart and Staveland, 1988), measuring perceived workload, was administered to subjects upon completion of each Audio type condition within an experimental session (Day 1 and Day 2). Subjects rated their workload in each Display x Audio experimental condition. A total of 6 sets of ratings were collected per subject.

Simulator sickness ratings. Participants' self reports of simulator sickness symptoms were collected at the end of each experimental session (Day 1 and Day 2) using the Kennedy Simulator Sickness Questionnaire (SSQ) (Kennedy & Lane, 1993). Baseline, pre-session symptom questionnaires were administered at the beginning of each experimental session for purpose of comparison.

Results

Objective Performance Measures

Separate 2 (Display type) x 3 (Audio type) x 3 (Block) x 2 (Trial) within-subjects repeated measures analyses of variance (ANOVAs) were conducted on speed and accuracy of target acquisition. Planned comparisons were examined on experimental variables of interest (e.g., Display, Audio) related to speed and accuracy performance independently.

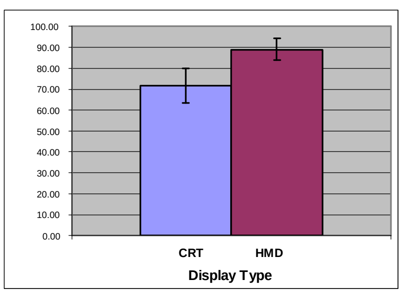


Figure 4. *Significant main effect of Display type.*

Targeting accuracy. A significant main effect of Display type was found in targeting accuracy, $F(1, 7) = 8.20$, $p < .05$ (Figure 4). Participants showed significantly more precise targeting when using a CRT ($M =$ 71.65 pixels) than an HMD ($M = 89.01$ pixels). No significance variance in performance was found as an effect of Audio type. No significant interaction of experimental variables was found. In sum, targeting was more precise when sensor feed was presented on a CRT than an HMD.

Targeting speed. A significant main effect of Audio type was found in speed of targeting acquisition, $F(2, 14) = 144.36$, $p < .001$ (Figure 5). Stereo and 3-D audio alerting ($M = 6.74$, $M = 7.06$; sec. respectively) supported more rapid target acquisition than Mono audio alerting ($M = 7.92$ sec.) No statistical difference in performance was found between Stereo and 3-D Audio conditions. No effect of Display type and no significant interactions were found. Overall, time on target was reduced with stereo and 3-D cues, when compared with performance with mono audio cues.

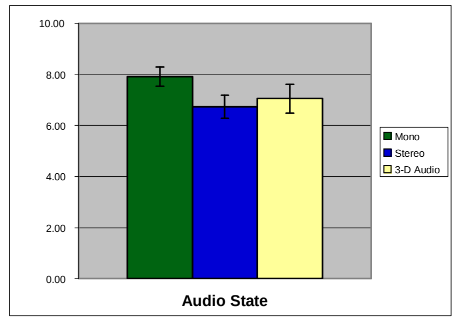


Figure 5. Main effect of Audio type ($n = 8$).

Missed targets. Data collected on the frequency of missed targets per mission showed a significant main effect of Audio type, $F(2, 14) = 5.84$, $p < .05$ (Figure 6). For missions where alerts were given in 3-D audio, participants were four times less likely to miss targets than when receiving mono audio alerts (.31:1.28 targets). Furthermore, 3-D audio alerts yielded an advantage of 2.7 times less missed targets than stereo alerting ($M = .84$). Participants showed significantly more missed targets in missions with mono audio alerts than all other audio conditions. No effect of Display type and no significant interactions were found. In summary, participants acquired the most targets when alerted in 3-D audio.

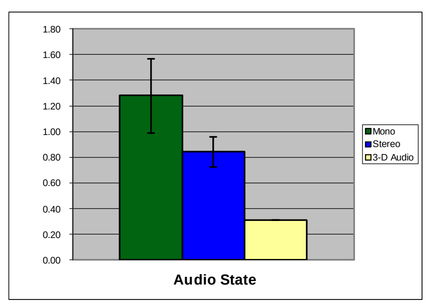


Figure 6. *Significant effect of Audio type (n = 8).*

Subjective Ratings

NASA-TLX workload ratings. In a comparison of means calculated from NASA-TLX ratings, collapsed across subscales, a significant interaction of Display and Audio type was found, $F(2, 14) = 4.26$, $p < .05$ (Figure 7). Missions flown using an HMD showed significantly lower operator workload when 3-D audio alerts were given. Whereas, both stereo and 3-D audio alerts positively impacted workload ratings when using a CRT. No significant difference in workload ratings was reported for mono and stereo alerting, when using an HMD. Data collapsed across Display type showed a significant main effect of Audio type (Figure 8), such that missions completed with 3-D audio alerts yielded significantly lower workload ratings than missions completed with mono or stereo alerting. No significant effect was revealed with the Display type manipulation. To summarize, missions flown with 3-D audio alerts yielded the lowest levels of reported operator workload. Furthermore, reportedly high workload ratings associated with the HMD were mitigated when 3-D audio cues were incorporated in the missions.

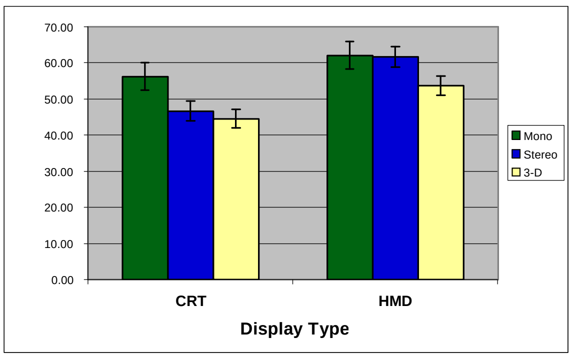


Figure 7. *NASA-TLX Workload Ratings show Display x Audio type interaction (n = 8).*

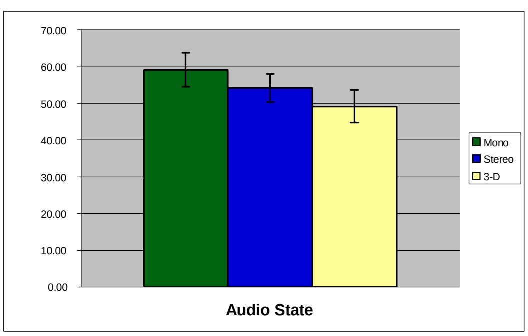


Figure 8. *NASA-TLX workload ratings show main effect of Audio State (n = 8).*

SSQ scores. Pre-session SSQ scores were calculated and analyzed for variance between Display types. As expected, no significant differences existed in reported sickness symptoms prior to exposure to the experimental session. Post-session SSQ scores, collapsed across sub-scales, revealed a significant interaction of Display type x Time, $F(1, 7) = 6.32$, $p < .05$ (Figure 9). Post experimental SSQ scores showed a significant increase in sickness symptoms when an HMD was used. Although, both sets of SSQ scores taken post session showed higher than baseline scores, use of the HMD showed sickness scores exceeding levels (SSQ > 20) warranted as tolerable by the developer of the questionnaire (Kennedy et al., 1992). In sum, participants reported more severe sickness symptoms with the HMD than the CRT, suggesting an unrecommended level of operator discomfort associated with HMD use.

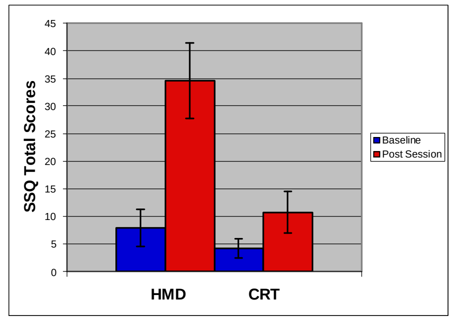


Figure 9. *Kennedy Simulator Sickness Questionnaire scores show Display type x Time interaction (n = 8).*

Discussion

In an evaluation of targeting performance data, accuracy was increased when the sensor feed was presented on a CRT versus an HMD. No associated performance tradeoff was recorded for time on target as an effect of Display type. It should be noted that when comparing CRT and HMD performance, not only the display, but the method of sensor control differed. Sensor control with the CRT was managed through fine motor input on a flybox joystick. By comparison, sensor control with the HMD was slaved to the swivel movement of a participant's head. Not unreasonably, precise targeting was better accomplished with fine motor movements of the practiced hand than more coarse movements of the head. In sum, the manipulation of display type revealed a performance decrement when the payload sensor was coupled to the head. Instead, results supported current joystick manipulation of the UAV sensor.

As anticipated, data collected on targeting performance revealed an effect of Audio alert type, supporting the use of stereo and 3-D audio alerts. These results upheld previous findings that directional audio cues reduce time to locate a target in a visual search task (Strybel & Guettler, 2001). For both stereo and 3-D audio alerts, participants were able to more rapidly acquire a target when given a directional audio cue, regardless of display type. Conversely, target search time was longer when the participants were given a non-directional (mono) cue. In an unexpected performance benefit, subjects given

3-D audio alerts missed four times less targets than when receiving mono alerts. Although it was hypothesized that 3-D audio alerting would enhance detection time; it was unforeseen that without 3-D localized alerting, subjects might miss up to 4 times more targets than non-directional alerting. Similarities in target search times for stereo and 3-D alerting conditions did not foretell the 2.7 times more missed targets for stereo versus 3-D audio. Therefore, it is important to consider both time on target and frequency of misses within the context of the operational scenario. In cases where rapid and successful acquisition of a high percentage of targets is necessary, performance data suggests the use of 3-D audio alerting.

The synergistic value of 3-D audio cueing paired with the HMD was revealed in reports of lower operator workload than experienced in all other audio conditions. Specifically, high workload ratings reported with HMD use were mitigated with 3-D audio cues. In both display types, 3-D audio cues supported lower levels of reported workload than alternate audio cues. As expected, operators experienced less workload when guided in a visual target search.

Reports of increased simulator sickness symptoms associated with HMD use and coupled sensor movement replicated findings of previous research (Morphew, Shively, & Casey, 2004). Hardware limitations of sensor slew rate and the associated visual lag likely contributed to self-reported nausea and eyestrain. As noted by the literature, even short periods of HMD use can result in side effects (e.g., headaches, nausea, blurred vision) that would be seen only after hours in front of a CRT (Stone, 1993).

Conclusion

Due to exceedingly high levels of reported sickness symptoms, head-slaved sensors with HMDs utilized in this study are not recommended. Once fatigue, stomach awareness, delayed sensor movement, and visual lag can be mitigated, HMDs may be a viable solution for UAV payload display and control.

Implications from this study suggest that 3-D audio alerting may offer enhanced capabilities to the payload operator for successful and rapid target acquisition. At present, 3-D alerting assumes target recognition technology that is not yet mature. Additional research will be required once developments of automated target recognition systems have reached operational proficiency.

Acknowledgements

Thanks to Daniel Casey of San Jose State University for his work in scenario development and Thomas Marlow of Silicon Valley Systems, Inc. for his technical assistance.

References

Begault, D. R., and Pittman, M.T. (1994). 3-D Audio Versus Head Down TCAS Displays. (Report No. 177636). NASA Ames Research Center.

de Vries, S.C., & Padmos, P. (1997). *Remotely controlled flying aided by a head-slaved camera and HMD* (Report TNO-TM 97-B024). Soesterberg, The Netherlands:TNO Human Factors Research Institute.

de Vries, S.C., & Padmos, P. (1998). Steering a simulated unmanned aerial vehicle using a head-slaved camera and HMD: effects of HMD quality, visible vehicle references and extended stereo cueing. *Proceedings of The International Society for Optical Engineering (SPIE) Conference*, (pp. 80-91).Orlando, FL.

DiZio, P., and Lackner, J.R. (1997). Motion sickness side effects and aftereffects of immersive virtual environments created with helmet-mounted displays. Presented at the RTO HFM Workshop, *The Capability of Virtual Reality to Meet Military Requirements,* Orlando:FL, December 1997. (RTO MP-54).

Draper, M. H., Ruff, H. A., Fontejon, J. V., and Napier, S. (2002). The effects of head coupled control and a head-mounted display (HMD) on large-area search tasks. Proceedings of the Human Factors and Ergonomics Society Conference.

Ericson, M. (2004). *3-D audio technology: 3-D audio displays provide military advantages.* Retrieved January 2005 from Airforce Research Lab Technology Horizons Web site: http://www.afrlhorizons.com/Briefs/Dec04/HE0402.html

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshakti (Eds.) Human Mental Workload, Amsterdam: North Holland Press.

Kennedy, R.S., & Lane, N.E (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology, 3(3)*, 203-220.

Kennedy, R.S., Lane, N.E., Lilienthal, M.G., & Hettinger, L.J. (1992). Profile analysis of simulator sickness symptoms: Application to

virtual environment systems. *Presence, 1 (3)*, 295-301.
Morphew, M. E., Shively, R. J., & Casey, D. (2004). Helmet mounted displays for unmanned aerial vehicle control. Proceedings of The International Society for Optical Engineering (SPIE) Conference, Orlando: FL.
Rash, C.E., & McLean, W.E. (1999). *Human factors and performance concerns for the design of helmet-mounted displays.* U.S. Army Aeromedical Research Laboratory: Ft. Rucker, Alabama. (USAARL Report No. 99-08)
Stone, B. (1993). Concerns raised about eye strain in VR systems. *Real Time Graphics, 2(4),* 1-13.
Strybel, T.Z., & Guettler, D.L. (2001). The effectiveness of visual vs. auditory auditory cues in a visual search performance: Implications for the design of virtual environments. In Smith, Slavendy, Harris & Koubek (Eds.), *Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality* (Vol 1) (pp. 1363-1367). Lawrence Erlbaum.
U.S. Department of the Army. (2001). *Technical Manual Operator's Manual for Shadow 200 TUAV System* (DTM 9-5895-681-10). Washington, D.C.

SAFETY STRATEGIES WHICH ALSO IMPROVE OPERATIONAL PERFORMANCE

Dr. Gary Eiff
Michael Suckow
Purdue University
West Lafayette, Indiana

Research has demonstrated that workers in aviation maintenance operations often perceive that safety and operational goals are in conflict. Investigators at Purdue University have worked with numerous aviation companies over the past eleven years to improve safety and control maintenance human errors. During that time, it has become apparent that safety goals, strategies and programs are differentially supported depending on the operational and economic pressures experienced by an organization. Purdue researchers have often traced operational and performance stressors back to poorly structured processes and other factors that result in artificially induced perceptions for the need to sacrifice safety for performance. Several strategies used or developed by Purdue researchers have demonstrated that safety and productivity gains can be simultaneously achieved through the use of process mapping and identifying areas in need of improvement.

Conflicting Goals

Studies by Purdue researchers at numerous aviation maintenance organizations have demonstrated that maintenance workers often feel that safety is compromised by work pressures and mixed messages from management. This phenomenon, commonly referred to as "conflicting goals" (Reason, 1997), is recognized by human factors researchers as a common cause for the erosion of operational safety levels and the diversion of worker focus from safety to productivity goals. In the maintenance organizations studied, there was strong support by management for safety in all facets of the operation and a stated mission of "safety first". How, then, does one explain the fact that the maintenance workers studied often felt that it was necessary to neglect safety procedures or shortcut safety steps in order to attain operational or productivity goals?

Purdue researchers found that in each of the organizations studied work related metrics for performance were one-sided. That is, they focused on the exclusive reporting of operational or productivity performance and failed to capture or report the level of safety of the operation (Eiff & Stanley, 2003). In short, work related performance metrics were providing feedback for productivity performance and not safety performance thus narrowing the perception of workers to a myopic view of what was important in work related outcomes.

Exasperating the impact of this imbalanced reporting of safety and productivity performance was the frequent occurrence of operational or productivity exigencies. Poor work plan development, process control and other operational factors often resulted in work related pressures which forced workers to feel they had to chose between doing the job safely and meeting operational or productivity goals. The subliminal message imparted to workers, as expressed to researchers by maintenance personnel, was that "safety is Number One unless it impacts operational performance or productivity".

In all of the organizations studied by Purdue researchers, overriding operational or productivity performance problems which led to this perceived pressure to sacrifice safety for other work goals were most often the result of poor process design and control, work coordination, and the failed understanding of how one's work performance impacted the overall productivity or operational performance of the organization.

Process Mapping Assessment Tool

Purdue researchers have repeatedly found that they have had to help organizations better understand how to analyze and improve their operational processes in order to improve workplace and operational safety and productivity. The strategy which has proven most effective at identifying, analyzing, and resolving operational problems has been the process mapping assessment approach.

When working with airline partner companies to identify and analyze operational problems, Purdue researchers generally begin by forming a group of company representatives to work together with the researchers on the project. These representatives are generally workers from each of the career fields affected by the problem. The initial phases of the project include providing training in process mapping and other techniques to be used in isolating and analyzing the problem. Once the whole project team is trained in the process analysis and improvement strategies to be employed, the team begins to define the operational process map by reviewing the airline's career field operation manuals to determine what the company's policies and procedures define as the company's approved way to perform the requisite tasks. This first draft of the process map normally results in the identification of policy and procedural inadequacies and the identification of many conflicts between the ways different operational manuals stipulate that identical processes should be performed. As a result, the team must begin its analysis process by resolving these procedural conflicts and revising the manuals to reflect one standard of operational performance.

Once this has been done, the team scrutinizes the map for unnecessary or redundant steps or processes. When the team has refined the map to the best of its ability, the map is then compared to how the process is actually done. Generally, the team finds that the process defined by the map is not the same as that being performed in the workplace. From this point on in the improvement cycle, the map is used to define operational process flow and feedback concerning the effectiveness of the defined process is used to improve the map and, eventually, the carrier operations manuals.

A Graphic View. Researchers have found that the highly graphic nature of the map makes it easily understood and usable by any worker. This causes the map to be the focal point of discussions between process improvement team members and workers or managers as they explore ways to streamline operational processes. The map has the additional advantage of providing workers at all levels of the operation a better understanding of operational work goals and the role they play in meeting those goals. It also provides them with an understanding of how they or their work group's tasks impact the overall operation. By utilizing the process map and following the process depicted, not only do the individual workgroups understand what is expected of them, but also the impact that their actions can have on members of other workgroups.

Task Coordination. One of the greatest contributors to the problems experienced by partnering aviation organizations was the coordination of workgroups' tasks and task integrations. Coordination of work tasks is made clear in the process maps through the alignment of the different task step lines for each workgroup along a common timeline. Therefore, if tasks are found to be occurring in parallel vertically, they are being carried out concurrently. Those that occur prior are located to the left, while those waiting yet to be completed are found to the right. In this way, individuals can look at the map to find out what should have already been completed both within their own group and by their peers in other groups to help them assess if the process is proceeding normally or if they should prepare their workers for likely deviations. In resolving study partner problems, researchers often guided the improvement team to a more effective process for workgroup coordination. This was often accomplished through the reduction of steps needed to accomplish the operational goal. By being able to assess the progress being made through the process, individuals can better plan how they will meet the needs of their own functions within the workflow in real-time.

The process map also provides graphic indication of important conjoining phases among work groups. Points within the operation which require the articulation or "hand-off" of tasks or completed processes between workgroups requires effective coordination and communication in order to insure undisrupted work flow. The process map makes the identification and analysis of both the timeliness and effectiveness of this coordination easier than with other methods.

Task Integration. Task integration is also portrayed well through the use of the process mapping technique. There is the perception among many workers that after they complete their parts of the operational process, they are no longer accountable for the success or failure of the process overall. However, as stated before, no one workgroup is able to complete their functions fully without both impacting and being impacted upon by the other workgroups involved. Therefore, the ability to interact in a professional and productive manner with members of different workgroups is a necessary part of any workflow. In the operational work environment of a station, very few tasks are accomplished exclusively by a single workgroup.

The process map expresses this notion by showing how the processes, as they are aligned, also require different steps to occur in sequences among workgroups. Using the maps as a tool, organizational members have the ability to follow the product through the process as it is moved along through the tasks of all the workgroups. Furthermore, the map can provide insights into locations within the map where integration and coordination are lacking so that the process can be improved and the map amended to reflect the new steps. Task integration was clearly demonstrated as a powerful result of the process mapping technique in the vast majority of industry problems addressed by the researchers. After developing and studying the operational process maps, the improvement teams found that rather than needing additional manpower and other resources, they were able to accomplish their goal by improving communication and coordination through the use of cell phones or by otherwise communicating with each other with key information at the predetermined critical junctures in the process. Through such usages of the workflow process mapping technique and strategic application of the communication-related insights gleaned from this tool, great successes in process improvement have been demonstrated in actual aviation operational settings.

Roles & Responsibilities. A major strength of the process mapping strategy is that it provides clarity of workgroup roles and responsibilities in a diagrammatically depicted representation of the progressive work process steps. This easy to understand perspective provides not only a "big picture" view of how the process strategically insures meeting organizational and operational goals but also provides adequate specificity to become a framework for tactical problem solving. Structured to map the flow of the product(s) through the organization's operational processes, this highly visual format aids in identifying and defining the process' critical path and subordinate critical chains. The process' critical path is the shortest series of necessary sequential steps required to meet the operational or productivity goals of the process. Critical chains are parallel work processes that must be integrated into the critical path at specific times during the process (Goldratt, 1997). The timeliness of the integration of critical chain products into the critical path is paramount to the successful completion of the operational objectives. After viewing the process map, workers from various workgroups clearly understand the role they play in meeting the organization's operational goals.

How Process Mapping Reduces "Risk"

Process mapping clearly assisted in the identification of roles and responsibilities, the reduction of operational process flow problems, and the coordination and integration of tasks in an operational setting but can it also be helpful in identifying "risk" in maintenance settings? As mentioned in the beginning of the article, the research literature suggests that workers engage in at-risk behaviors when work pressures make them feel that they must sacrifice safety for productivity and operational goals (Reason, 1997). The process mapping strategy provides for an easy assessment of impediments to effective and efficient workflow that cause work disruption or pressures that result in worker at-risk behaviors. Purdue researchers have found positive correlations between the use of process mapping and the identification of work practices that may include unsafe operating practices or unauthorized work practices that may improve efficiency in the short run but are clearly in the high risk category of application. It also helps identify incorrect or inadequate policies, procedures, or work habits. These facets of the work process are crucial to establishing worker behaviors as they serve as the antecedents for worker behavior (Braksick, 2000). Correct worker antecedents are a pivotal step in correcting unwanted behaviors that impact safety and productivity.

The process map also identifies where, when, and between what work groups critical task coordination or integration occurs. These conjoining points most often represent interfaces between "critical chain" and "critical path" processes. It is precisely at these points that many operational problems and workflow delays occur and that increased safety exposure is generated (Eiff & Lopp, 2001). By focusing on more effective communication and work coordination at these points, researchers have been able to improve workflow and, at the same time, reduce risks which have previously resulted in accidents or safety incidents.

Improved Use of Resources

Another factor which exacerbated problems, especially at conjoining points, was the lack of adequate resources to perform the tasks of the process. At all of the organizations studied, Purdue researchers found that at critical work "turn-over" points in the process, problems often were generated or compounded by the lack of necessary resources for the effective completion of tasks (Eiff & Lopp, 2001). Resource deficits often include fewer than required workers to perform the task effectively, inadequate equipment resources, or missing supervision. Building on the foundation of process map analysis technique, Purdue researchers used the operational maps to develop a resource assessment and utilization strategy which provided organizations with insight into adequate staffing and resource levels which allowed for optimization of resources. This assessment strategy has been utilized to develop manpower and equipment resource planning guides to aid managers in allocating appropriate resources to accomplish tasks effectively and efficiently. The tool also provides managers with insight into changing resource needs in the constantly changing operational environments normally associated with airline operations.

Systems Thinking

The highly understandable process mapping strategy also helped researchers explain to managers and workers in the studied organizations the need to address work group isolationism. It is common among aviation organizations for the workforce to become "soiled" in their own professional work groups or environments. When this occurs, workers often fail to see how their work performance or safety focus can impact other work groups or the organization as a whole. Effective safety and performance gains can be realized if the organization management and workers can take a more global or systems view of their operation. Thus, moving the organization toward "systems thinking" can have a dramatic impact on both safety and operational performance.

Systems' thinking was an additional methodology used by researchers to assist in understanding the holistic perspective within organizational settings and the perceived conflicts between operational and safety goals. Principles of system thinking include:

- Think of the "big Picture"
- Balance shot-term and long-term perspectives
- Recognize the dynamic, complex, and interdependent nature of systems
- Take into account both measurable and non-measurable factors
- Interrelatedness of systems
-

(Anderson & Johnson, 1997)

Process mapping allows researchers to better visualize and inform others of the work flow and identify limitations with a view toward reducing the scope of the work involved to the simplest and smallest steps. System thinking expands the vision to include multiple systems and how the dynamics involved may have unanticipated outcomes by virtue of the interrelatedness of all the subsystems and, while they can be analyzed in isolation, they cannot be solved without taking into consideration emergent effects in other areas of the organization.

Three Solution Categories; Personnel, System, Documentation

The use of these various tools in unison represents a more balanced approach to resolving troublesome workflow and, therefore, safety and productivity problems. Once the tools identify operational risks and performance impediments, three categories of need must be considered when revising operating practices. The first is personnel issues. Are there enough personnel to perform the tasks with optimal performance? Do personnel have sufficient knowledge, skills and abilities to complete the tasks as assigned in a manner consistent with the new information or should they be trained? Is the operation function in accordance with a systems approach to operational goals? Or, do job tasks need to be redesigned to insure systems compatibility? Is the process resourced adequately throughout the workflow, is the tooling available as required, is there an adequate support system for employees to receive appropriate timely feedback. Another area of concern is that of workflow and task documentation. Documentation on work process flow which specifically addresses the tasks and performance criteria for the operation provides the important antecedents for correct worker performance. These well defined antecedents are the precursors to better productivity and safety performance. It is also true that good audit trails rely on adequate documentation and are a critical component of internal and external checks and balances. A sound risk management program relies on good documentation and accurate data collection systems.

Summary

The highly intuitive nature of the process mapping technique has many advantages. It is easy for workers, managers, and researchers to use in the identification, analysis, and improvement of operational and process problems which often drive safety concerns. It has been demonstrated that the process mapping technique is also highly effective at providing insight into critical points in the process where safety problems arise and for determining the root causes for those problems. Spin-off techniques such as task coordination and integration, resource utilization, and system structure and thinking analysis and improvement strategies have proven to be dramatic enhancements to the fundamental technique of process mapping. Together, these strategies have demonstrated a highly effective way to improve both safety and operational productivity simultaneously. In today's troubled industry, such tools could prove pivotal for organizations with bleak economic outlooks.

References

Armentrout-Brazee, C. & Eiff, G. (2002, March 3-5). Communication in process: improving communication through the analysis of airline operational processes (pp. 1-8), *Aviation Communication Conference*, Phoenix, AZ.

Braksick, L. W. (2000). *Unlock behavior, Unleash profits.* New York: McGraw-Hill.

Eiff, G. & Lopp, D. (2001, May 14-16). Measuring and correcting workers at-risk behaviors. *Proceedings of the Canadian Aviation Safety Symposium* (pp. 1-17), Transport Canada, Vancouver, British Columbia, Canada.

Eiff, G., & Stanley, D. (2003, July 21 &22). The problem with measuring operational safety, Aviation Management Education and Research Conference (Group 2 (2), pp. 1-17). Montreal, Quebec, Canada.

Goldratt, E. M. (1997). *Critical chain*. Great Barrington, MA: The North River Press.

Reason, J. (1990). *Human Error*. Cambridge, England: Cambridge University Press.

Reason, J. (1997). *Managing the risks of organizational accidents*. Brookfield, Vermont: Ashgate Publishing Co.

Anderson, V. & Johnson, J. (1997). *Systems Thinking Basics: From Concepts to Casual Loops.* Williston, Vermont: Pegasus Communications, Inc.

ENGLISH AS WORKING LANGUAGE FOR NON-NATIVE SPEAKERS–THE ASSESSEMENT OF ENGLISH LANGUAGE PROFICIENCY AMONG AB-INITIO APPLICANTS

Hinnerk Eißfeldt

Deutsches Zentrum für Luft- und Raumfahrt DLR
(German Aerospace Center)
Institute of aviation and space medicine
Department of aviation and space psychology
Hamburg, Germany

In aviation English has been agreed upon being the international working language ever since. However only less than 15% of the worlds population speaks English as mother tongue, and it seems reasonable to assume that among pilots and controllers the percentage of native speakers is below 30%. To secure high global standards the International Civil Aviation Organisation ICAO has recently defined new requirements concerning the level of English language proficiency among aviation professionals. From March 2008 on aviation professionals have to be assessed concerning their proficiency in speaking and listening preferably in aviation-specific context. ICAO proposes to start formal evaluation much earlier to assure applicants to meet language proficiency requirements as a prerequisite for recruitment. However by now no validated tools to achieve this have been published. This article offers a solution derived from the experience of the German Aerospace Center DLR to test English language skills among applicants for aviation careers, for example pilots, air traffic controllers or even astronauts.

ICAO Language Proficiency Requirements

In 1951 the International Civil Aviation Organisation ICAO reached a decision supported by all member states that “pending the development and adoption of a more suitable form of speech for universal use in aeronautical radiotelephony communications, the English language should be used as such and should be available on request” (ICAO recommendation 5.2.1.1.2). Detailed phraseology was developed thereafter to avoid miscommunication between partners in radio communication. However this did not prevent communication to play a significant role in incidents or accidents (for a listing see Jones 2003). Tenerife in 1977 (583 ...), Avianca052 in 1990 (73...) are the most prominent examples for the deadliness of deficient language skills in aviation. According to ICAO “between 1976 and 2000 more than 1.100 passengers and crew lost their lives in accidents in which investigators determined that language had played a contributory role” (Mathews 2004). Detailed safety analyses have revealed that the proper use of predefined ATC phraseology is not always sufficient. Thus in 2003 ICAO has released amendments to annexes of its Chicago Convention requiring aviation professionals involved in international operations to demonstrate a certain level of English language proficiency. As ICAO now states in special circumstances pilots and controllers must be able to express themselves in plain language.

Annex 10 describes what language(s) shall be used for radiotelephony communication: the language of the ground station OR English. This means that proficiency in ICAO phraseology and plain English is required. Annex 6 and 11 establish that all personnel (pilots and air traffic controllers) comply with the ICAO language proficiency requirements stipulated in Annex 1. Annex 1 describes the language proficiency and testing requirements and contains a rating scale with six proficiency levels. Table 1 lists the proficiency levels defined by ICAO and the amount of retesting necessary.

Table 1. *English language proficiency levels defined by ICAO*

Level 6 (Expert)	will not be required to demonstrate subsequent language proficiency.
Level 5 (Extended)	will need to be retested every six years.
Level 4 (Operational)	will need to be retested every three years.
Level 3 (Pre-operational)	or below:
Level 2 (Elementary)	will need specific Aviation English language training
Level 1 (Pre-elementary)	to reach the minimum ICAO Operational level.

The minimum language proficiency is defined at ICAO Level 4 (Operational) as a licensing requirement. Table 2 describes the rating scale at this level. Although these standards became applicable in November 2003, all ICAO Member States have been given until March 2008 to fulfill the necessary training requirements to allow personnel to meet mandatory testing and licensing requirements. States not incompliance with the new licensing requirements will be requested to notify ICAO, which may limit international recognition of licenses.

Table 2. *ICAO language proficiency rating scale (Operational Level 4)*

ICAO language proficiency rating scale (Operational Level 4)	
Pronunciation *	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation but only sometimes interfere with ease of understanding *Assumes a dialect and/or accent intelligible to the aeronautical community
Structure *	Basic grammatical structures and sentence patterns are used creatively and are usually well controlled. Errors may occur, particularly in unusual or unexpected circumstances, but rarely interfere with meaning *Relevant grammatical structures and sentence patterns are determined by language functions appropriate to the task
Vocabulary	Vocabulary range and accuracy are usually sufficient to communicate effectively on common, concrete, and work-related topics. Can often paraphrase successfully when lacking vocabulary in unusual or unexpected circumstances.
Fluency	Produces stretches of language at an appropriate tempo. There may be occasional loss of fluency on transition from rehearsed or formulaic speech to spontaneous interaction, but this does not prevent effective communication. Can make limited use of discourse markers or connectors. Fillers are not distracting.
Comprehension	Comprehension is mostly accurate on common, concrete, and work-related topics when accent or variety used is sufficiently intelligible for an international community of users. When the speaker is confronted with a linguistic or situational complication or an unexpected turn of events, comprehension may be slower or require clarification strategies.
Interactions	Responses are usually immediate, appropriate, and informative. Initiates and maintains exchanges even when dealing with an unexpected turn of events. Deals adequately with apparent misunderstandings by checking, confirming, or clarifying.

Testing of English Language Skills at DLR

English language testing has always been part of DLR's test system. A standard test battery for pilots or air traffic controllers for example to our mind has to contain a written test of English (grammar, vocabulary, meaning) in a multiple-choice format to be applied in groups of up to 50 candidates in the first stage of selection (a more detailed description of the selection system is provided by Eißfeldt & Deuchert 2002). Under special circumstances even more than one test has to be used at this stage to include an early assessment of the ability to understand spoken information For candidates reaching the second stage of selection their actual English skills has to be assessed on an individual base either in a special oral examination or during the interview (e.g. if this is to be done in English anyhow). If the candidate applies for a job in a multinational team with English being the working language, also native speakers shall be assessed regarding language skills as the intelligibility of their voice output might be restricted due to strong dialect. Problems of dialect and pronunciation are also reasons why ICAO demands aviation professionals to be assessed in their national language too. With the new ICAO requirements for training providers it will be very important to assess the proper level of English language prior to the start of training, as according to the new regulations insufficient language skills will terminate training of any applicant regardless of all other achievements. In the following it is described how English language proficiency can be assessed among ab-initio applicants using existing DLR tests.

English Listening Test ENL

The English Listening Test" ENL" was developed in 1993, when the German Aerospace Center DLR was in charge of the selection of international air traffic controller applicants for EUROCONTROL. At that time tests in use concerning English language skills used either written items of multiple-choice format or spoken English items, for instances vocabulary that had to be translated in writing or numbers that had to be written down. This required a lot of manpower as it did not allow for machine based scoring techniques. In addition after seeing applicants in the interview the impression occured that although test scores have been at level for some applicants the language competence to conduct an interview in plain English was rather restricted. To avoid a waste of time in the selection process the newly developed test should measure the understanding of complex meaning on the basis of spoken English language and allow for machine scored group testing.

The test offers pure acoustic items presented via headset to workon. Some of the items refer to aviation to increase the applicant`s motivation. The language used is exclusively British English. To control the impact of mother tongue in the sample, all steps of test development were performed twice, including or excluding native speakers. The test consists of four different parts. All parts require to listen to acoustic information first. Then four alternative are presented to choose the correct answer. The time to choose one of the four answers is restricted.

Each of the four parts of the test assesses English listening comprehension in a different format. The four parts are:

Part 1 - Simple Meaning (12 Items), where a sentence is read and the test taker needs to find outwhich of the four given options presents the sentence that is closest in meaning to the one heard;

Part 2 – Numbers (10 Items) where a sentence including a number is read and the test taker has to choose the number mentioned in the sentence from four answers offered;

Part 3 – Vocabulary (12 Items) where a sentence is read and one of the words is marked. The test taker has to choose out of four options a word that is closest in meaning to a certain word that was read in the sentence.

Part 4 - Complex Meaning (12 Items) where a short story of about 100 words is read and questions relating to the story are presented.

The test administration itself is fully computerised. The test taker has to click with the mouse onto the frame that contains the correct response or put a finger on the touchscreen accordingly. A test administrator is needed in order to introduce the test taker and to monitor the testing process. In particular, disturbing noise has to be prevented and it is not allowed to take notes during the test or to refer to dictionaries. The scoring procedures are fully computerised and the test is evaluated automatically. In a special application the ENL is administered and evaluated via internet.

ENL results are reliabel: Cronbach's α for the computerised test version of the test was 0.89 (n=194) in a study conducted with European ATC applicants in 2000. Construct validity is proven by the correlation of of the ENL total score with the result of a written English test (ENS, English written) with r=0.80, p<.000, n=403. After exclusion of native speakers (Origin: Great Britain) the correlation was r=0.76, p<.000, n=341. ENL and ENS were both administered at the same testing session (pre-selection stage) at different times of the day.

To assess predictive validity ENL test results were used to predict results of English oral examination, which was done several weeks after the first stage. At the end of the second testing stage (main selection) an oral interview was conducted by the interview board with those applicants having passed all other tests. Directly after the interview, five selection board members rated the applicants' oral performance in English in a quasi-Stanine scale. The average of those ratings forms the final score for oral English (ENM). The correlation of ENL total score and ENM was r=0.69, p<.000, n=109. Excluding native speakers (origin: Great Britain) the correlation of ENL with ENM was r=0.66, p<.000, n=93 in a sample comprising 21 different European nations.

Standard Oral Examination

The standard oral examination at DLR is developed for non-native speakers. It is performed in a standardized manner using special item material and defined measurements. The candidates have 15 minutes to read a text of about one page lenghth to prepare for the examination. They then have to read it aloud in front of the board, retell the story in their own words and answer some questions. In the second part candidates are free to choose among different types of items: pictures, cartoons (picture stories) or general statements to be used as basis for interaction in free speech.

Usually the oral examination is performed by job incumbents after having received a special training as for instance in the selection of ab-initio air traffic controllers for Deutsche Flugsicherung GmbH DFS. Criteria to be rated are pronunciation, grammar, vocabulary and comprehension. Every criterion is described by 3-4 anchored subscales on a standard rating form. As Stanine scales are used throughout the selection process,the overall rating as well as the criteria are measured on a quasi-stanine scale. Interrater correlations rank from r= .72 to r=.85 for the criteria and r=.89 (all p<.000, N = 660) for the overall

English oral stanine score. Figure 1 shows the distribution of results for N = 660 candidates.

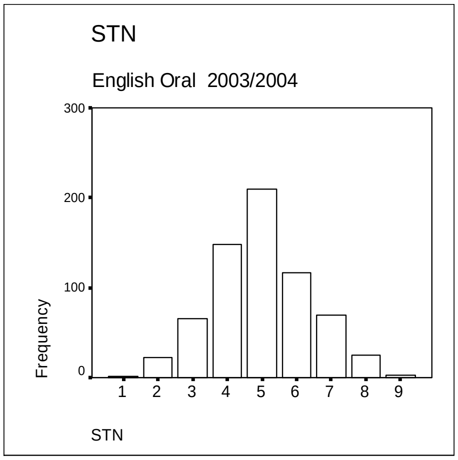


Figure 1. *Results of English oral examination, N = 660*

In the context of selection of controllers for DFS the English oral examination is of special importance as for candidates receiving a result just below the required level a special option is available. Provided that all other test results (cognitive testing, work sample tests, assessment center and interview) are at or above the defined level of acceptenceand the candidate would be recommended for training course elsewise, he/she can retake the English oral after some additional training of within half a year. It then depends on the initiative of the candidate to improve his/her English on his own costs. About 80% of candidates retaking the English oral are finally successful and enter ATC training. Their success rate in institutional as well as in practical training is the same compared to trainees without special additional language course.

English Language Competence and Training Success

The predictive power of English language test performance has been assessed in different validsdation studies at DLR. Usually test results in English show close correlation not only with English grades at school but with school grades in general. In a detailed study the general mental ability 'g' was computed for N=2954 air traffic control applicants using the various test results in selection (see Damitz & Eißfeldt 2004 for details). When 'g' was correlated with the results from each single test, results indicated a strong connection between 'general mental ability' and foreign language skill ($r=.40$, $p<.000$). Furthermore in a national validation study with ATC trainees English appeared to be among the best predictors of theoretical training at the academy as well of the simulator checks (Damitz et al 2000). Although some of the content of training is presented in English strong correlations have also been found for examinations not related to foreign language. Similar findings occurred in a validation study with ab-initio pilots in Asia. Thus a solid level of English language proficiency as it is required in ICAO level 4 will not only increase aviation safety but also has the potential to reduce failure rates in training among ab-initios. Using the proposed DLR tests can be of great help assessing English language proficiency as they are easy to administer and have been succesfully applied in aviation for many years. Providing norms reflecting international samples can be of major advantage when ICAO intends to guarantee the same language criteria to be used across all Member States.

References

Damitz, M. & Eißfeldt, H. (2004). The relevance of general cognitive ability (g) for training success of ab-initio air traffic controllers. In K.- M. Goeters (Ed.), *Aviation Psychology: Practice and Research* (pp. 135-140). Aldershot: Ashgate.

Damitz, M., Eißfeldt, H., Grasshoff, D., Lorenz, B., Pecena, Y. & Schwert, T. (2000). Validierung des DLR-Auswahlverfahrens für Nachwuchs-fluglotsen der DFS Deutsche Flugsicherung GmbH: Ergebnisse des Projektes Qualitätssicherung. In *DLR FB 2000-45*. Hamburg: DLR.

Eißfeldt, H. & Deuchert, I. (2002). The Selection System in use in Germany. In H. Eißfeldt, M. C. Heil & D. Broach (Eds.), *Staffing the ATM system - the selection of air traffic controllers* (pp. 121-130). Aldershot: Ashgate.

Jones, R. K. (2003). Miscommunication between pilots and air traffic control. In *Language Problems & Language Planning, 17, 233-248*.

Mathews, E. (2004). New provisions for English language proficiency are expected to improve aviation safety. *ICAO Journal*, *59*, 4-6.

FLIGHT SIMULATION AS AN INVESTIGATIVE TOOL FOR UNDERSTANDING HUMAN FACTORS IN AVIATION ACCIDENTS

Bart Elias
The Library of Congress, Congressional Research Service
Washington, DC

As aviation accidents become more complex, the need to better understand human factors aspects increasingly requires more sophisticated investigative tools and techniques. The capabilities of modern research, engineering, and flight training simulators provide human performance investigators with unique opportunities to reconstruct aviation accidents and test hypotheses regarding possible causal and contributing human factors. Two recent examples - the investigation of a flight test accident in October 2000, and the investigation of the crash of American Airlines flight 587 in November 2001 - highlight the potential role of flight simulation as an investigative tool. Building on these experiences, agencies responsible for accident investigations may consider adopting more formal guidelines for using flight simulation as a tool for understanding human factors in aviation accidents.

Background

Increasingly, human performance investigators are turning to simulation to reconstruct complex accidents and test specific hypotheses regarding possible human factors aspects. The improved fidelity and enhanced capabilities of modern research, engineering, and training simulators provide unique opportunities to reconstruct aviation accidents and test hypotheses regarding possible causal and contributing human factors. Using flight simulation, members of an investigation team can experience an accident reconstruction in an immersive environment that provides great flexibility to examine the accident sequence as a whole or focus on specific details.

Experiencing an accident reconstruction is, of course, highly dependent on one's background, training, and perspective. For example, a pilot is likely to focus on very different features than an aircraft performance specialist or a simulator engineer. Similarly, an airline or pilot union representative is likely to have a very different perspective than an airline manufacturer's representative that they bring to bear when assessing investigative activities conducted in a simulator. Therefore, an emerging role for human performance investigators is to lead multidisciplinary teams through these simulator reconstructions and synthesize the unique perspectives of team members into a cohesive understanding of the reconstructed accident sequence. Another emerging role for human performance investigators is to develop and test specific hypotheses regarding possible causal or contributing human factors using flight simulation.

Case Studies

Both of these roles were exemplified in two recent National Transportation Safety Board (NTSB) investigations that highlight the potential role of flight simulation as an investigative tool for understanding human factors in aviation accidents. During the investigation of a recent flight test accident in Wichita, Kansas, the investigation team used both an engineering simulator and a motion-based training simulator to reconstruct the accident and test hypotheses regarding pilot performance. Also, during the investigation of the American Airlines Flight 587 accident, the NTSB's human performance group participated in an accident reconstruction using a unique research simulator, the NASA Ames Vertical Motion Simulator (VMS), and examined upset recovery training procedures in the airline's training simulator.

Bombardier Challenger Flight Test Accident

Accident Summary. On October 10, 2000, a Canadair Challenger, operated by Bombardier Incorporated, crashed during initial climb from the Wichita Mid-Continent Airport (ICT), Kansas. The airplane was departing for a test flight to evaluate stick force characteristics of a new pitch-feel system (PFS) installed for European certification requirements. The aircraft's center of gravity (CG) had been set at the certified aft limit for the purposes of the flight test. The pilot and flight test engineer were fatally injured in the crash. The copilot was seriously injured and later died from his injuries. The NTSB determined that the probable cause of the accident was the pilot's excessive takeoff rotation combined with a rearward shift in the airplane's CG due to fuel migration that placed the airplane in a stall at too low of an altitude for recovery (NTSB, 2004, April 14).

Simulator Studies. A central issue in the investigation was the possible contribution and potential interactions between the pilot's takeoff technique, the flight test's aft-CG configuration, and the PFS installed for testing. To examine these issues in detail, the investigative team conducted studies using an engineering simulator and a level-D training simulator.

Two studies were conducted in the Bombardier Aerospace reconfigurable engineering flight simulator (REFS). In both studies, two type-rated pilots flew takeoff runs in the REFS which was configured with engineering models of the Challenger aircraft. At the completion of each takeoff run, the pilots provided a Cooper-Harper rating assessing the airplane handling characteristics with respect to pitch force, rotation rate, and ability to capture and hold the target pitch attitude (see Cooper and Harper, 1969). The pilots also provided comments on airplane handling characteristics.

The first study was conducted to assess the effects of airplane CG location on rotation rate and the ability to capture the target pitch attitude (nominally, 14 degrees). After completing several "familiarization flight" takeoffs performed with the CG set at 37.9 percent mean aerodynamic chord (% MAC), the pilots completed several takeoffs with the airplane's CG set at six different locations between 35% and 42% MAC. While the NTSB calculated the static CG for the accident flight to be 37.9% MAC and the airplane's certified aft-CG limit is 38% MAC, four CG positions aft of this limit were included in the test because the NTSB determined that a rearward fuel migration shifted the CG on the accident flight to about 40.5% MAC.

The CG locations were randomly presented, and the pilots were not told what CG location to expect for a given takeoff. Each pilot completed two takeoffs at each CG location using normal rotation techniques to achieve a rotation rate of about 3 deg./s. Each of the six CG locations was presented once before being repeated in another randomly ordered set of six takeoffs. Once the pilots had completed these 12 takeoffs using normal rotation techniques, they completed another two takeoffs at each CG location. However, for these takeoffs, the pilots were instructed to use a more "aggressive" rotation technique to try to achieve a rotation rate of about 6 deg./s.

Figure 1 presents the mean Cooper-Harper ratings for each combination of CG location and rotation technique. The pilots generally gave significantly higher Cooper-Harper ratings, indicating more handling difficulties, when they were instructed to use increased rotation rates. CG location, on the other hand, had no statistically significant effect on the Cooper-Harper ratings. However, the pilots, in general, commented that forward CG positions caused the simulator control column to feel heavy, while aft CG positions caused them to rotate at a somewhat higher rate and overshoot target pitch attitude slightly. The pilots generally noted that these effects were more noticeable when they used increased rotation rates. When increased rotation rates were used, the pilots noted that the stick shaker frequently activated, but usually for only short periods of time. The pilots also indicated that the simulator was controllable at all CG locations using both normal and increased rotation rates.

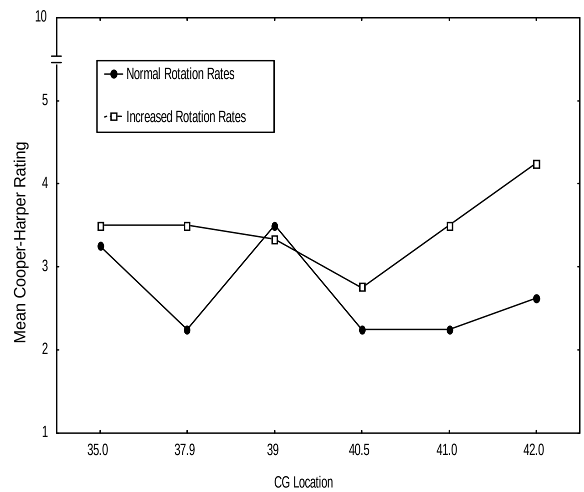


Figure 1. *Mean Cooper-Harper Ratings as a function of CG location and rotation technique.*

A second study was conducted in the REFS to assess any perceptible differences between the handling characteristics of the PFS installed for the flight test in the accident airplane and the PFS installed on certified Challenger airplanes at the time of the accident. Each pilot performed takeoffs with either the modified or production PFS units and provided Cooper-Harper ratings and comments. The CG was set at 40.5% MAC for each takeoff. The pilots reported no handling differences between the modified PFS and the production PFS and there were no significant differences in Cooper-Harper ratings.

Studies were also conducted in a motion-based training simulator. A reconstruction of the accident was played in the training simulator to allow investigators to experience the accident sequence. Also, a scaled-down version of the rotation rate study performed in the REFS was conducted in the training simulator. However, the most notable study conducted in the training simulator was designed to examine the perceptibility of rotation rates. At issue was whether the accident pilot's rotation technique was noticeably more aggressive than recommended procedures dictate during flights prior to the accident. The motion-based simulator was back driven with recorded flight test data from four Challenger takeoffs flown by the accident pilot and one comparison takeoff flown by another Bombardier flight test pilot. The peak pitch rates of the sample takeoffs flown by the accident pilot were 4, 6, 6.5 and 7 deg./s while the peak pitch rate of the comparison takeoff was 3 deg./s. Three pilots took turns experiencing the takeoffs from both the left and right seats of the simulator. The right seat occupant performed routine pilot not flying (PNF) duties during each takeoff while the left seat occupant manually followed the back-driven controls throughout each takeoff. On each run, the pilots observed two takeoffs, the comparison takeoff and one of the four sample takeoffs flown by the accident pilot. The comparison takeoff was presented either before or after the sample takeoff and the participants were not told what takeoffs they would be observing on any given run. After each run, the pilots compared the perceived peak rotation rate of the two takeoffs.

The data were analyzed by comparing the estimated (perceived) difference in peak rotation rates between the sample takeoff flown by the accident pilot and the comparison takeoff to the actual difference. The mean differences between estimated and actual peak rotation rates for each of accident pilot's takeoffs observed are shown in *Figure 2*. If the pilots had accurately estimated the rotation rates, then these values would be zero. Positive values reflect an overestimation of pitch rate while negative values indicate an underestimation. In general, there was a slight tendency among the pilots to overestimate the peak rotation rate for the flight with the lowest peak rotation rate and underestimate the peak rotation rate for flights with faster rotation rates. This trend was somewhat more pronounced for pilots occupying the right seat and performing PNF duties, but these differences in ratings between left and right seat observers were not statistically significant.

As evidenced from these studies, flight simulation was used extensively to address specific human performance questions regarding the handling characteristics of the airplane and pilot rotation technique. Human factors studies were conducted using flight simulators to test hypotheses and support conclusions regarding specific human factors aspects of this accident. Flight simulation was also used extensively to study pilot actions and flight control systems characteristics in the crash of American Airlines Flight 587.

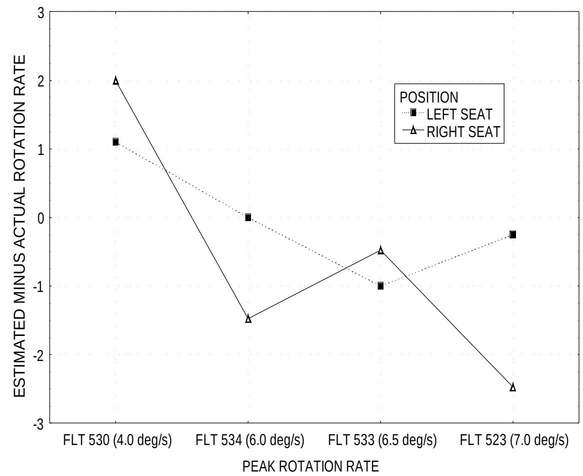


Figure 2. *Estimated minus actual peak rotation rates derived from the rotation rate comparison data*.

American Airlines Flight 587

Accident Summary. On the morning of November 12, 2001, American Airlines flight 587, an Airbus A300-600, was destroyed when it crashed into a residential area shortly after takeoff from the John F. Kennedy International Airport (JFK). All 260 on board and 5 people on the ground were killed in one of the deadliest crashes in U.S. history. The NTSB determined that the airplane crashed following an in-flight separation of the vertical stabilizer caused by excessive and unnecessary rudder pedal inputs. The NTSB also found that the rudder system design and the techniques used to train the pilot in upset recovery were contributing factors (NTSB, 2004). Human factors were a central focus of this investigation, and flight simulation proved to be an important tool in studying these factors.

Simulator Studies. To reconstruct the accident sequence and examine the acceleration forces and motions experienced by the pilots preceding the accident, the NTSB conducted tests and observations at the NASA Vertical Motion Simulator (VMS), a unique facility located at Moffett Field, CA (NTSB, 2002, October 3). The VMS, depicted in *Figure 3*, offers unparalleled capabilities for replicating large amplitude motion cues. The VMS cab is mounted on a six-degree-of-freedom motion platform that provides the following motion capabilities, making it the world's largest motion based simulator:

Table 1. *VMS Nominal Motion Limits*

Motion	Range	Velocity	Acceleration
Vertical	±30 ft	16 ft/sec	24 ft/sec/sec
Lateral	±20 ft	8 ft/sec	16 ft/sec/sec
Longitudinal	±4 ft	4 ft/sec	10 ft/sec/sec
Roll	±18 deg	40 deg/sec	115 deg/s^2
Pitch	±18 deg	40 deg/sec	115 deg/s^2
Yaw	±24 deg	46 deg/sec	115 deg/s^2

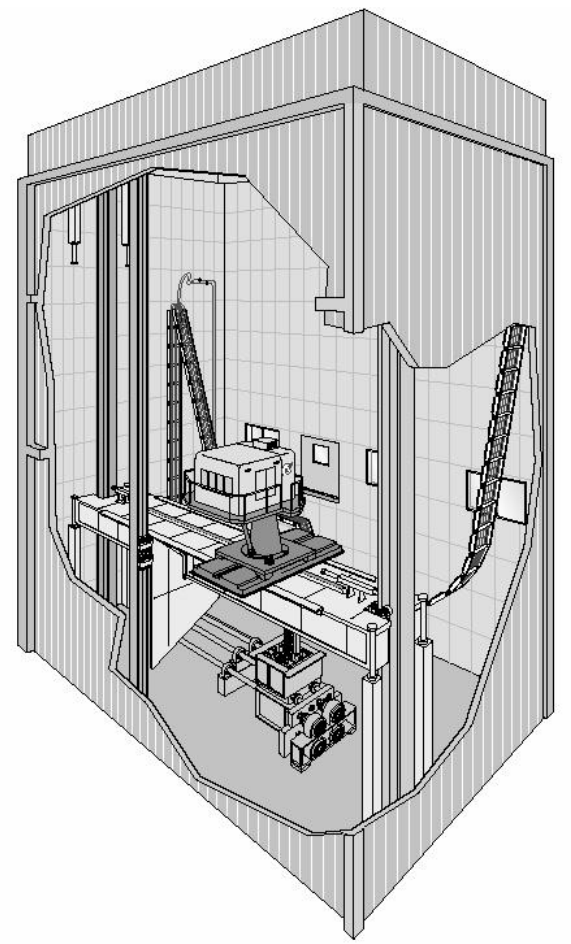


Figure 3. *A cutaway view of the NASA VMS facility (NASA Ames Research Center, 2000, October).*

The accident was reconstructed in the VMS using data derived from the accident aircraft's digital flight data recorder (DFDR) and calculations made by NTSB aircraft performance specialists. Audio segments of the accident aircraft's cockpit voice recorder (CVR) were synchronized for playback during the simulations. The VMS simulator cab was configured with two side-by-side pilot stations, each equipped with three side-by-side CRT monitors. At each station, the outboard monitor presented graphical strip charts of the input (derived from DFDR data) and actual (simulator cab recorded) accelerations for pitch, roll, and yaw axes, and flight control positions. The inboard monitor displayed a compass rose navigation display and the center monitor presented a primary flight display (PFD) similar to those in Airbus A300-600 airplanes.

The cab motion, flight displays, primary flight controls –including the rudder pedals, control wheel, control column–and throttles were back-driven from interpolated DFDR data during simulator runs. Although the cab was equipped with gear, spoiler, and flaps levers, these controls were not back-driven during simulator sessions. Some observers elected to manipulate these controls in response to CVR events to further involve themselves in the accident reconstruction. An out-the-window visual scene of prominent visual features and coastline in the vicinity of JFK airport was presented during simulator sessions. The simulator sessions were videotaped and a cockpit push-to-record button allowed participants to record verbal comments for later reference.

After the nine member human performance group spent two days experiencing the accident reconstruction in the VMS, the group met to formalize their consensus observations. The group focused on the accelerations and motions produced by encounters with wake turbulence and the subsequent flight control inputs made by the first officer on the accident flight. Many of the group members described the first encounter with wake turbulence as typical of a crossing wake encounter. Some participants felt a slight yaw before the flight controls moved. The slight yaw was described as a characteristic motion of an A300 flying through turbulence. This was followed by a vertical acceleration, described by the participants as a "bump", that seemed to result from the wake encounter rather than flight control movements. No flight control inputs followed this event. The group members generally agreed that "very slight" cab motions were felt as a result of a second wake turbulence encounter a few moments later that immediately preceded the initial movements of the control wheel and rudder pedal to the right. The cab motions were described as "barely perceptible" left lateral accelerations. Most participants did not experience any cab motion until less than one second before the first wheel motion. The first movements of the control wheel and rudder pedal to the right were considered to be "large and abrupt." The participants did not observe a visual or acceleration cue that would cause a pilot to apply the magnitude of wheel and pedal inputs observed. Transport pilots in the group noted that the large magnitude and rapid speed of these inputs were analogous to potential flight control inputs made during an avoidance maneuver. After these first movements of the wheel and pedal to the right, large lateral accelerations were felt, and additional large, abrupt flight control movements in the yaw, pitch, and roll axes were observed.

While the Airbus A300-600 is equipped with a variable stop rudder system that limits rudder pedal travel at higher airspeeds, the VMS was also used to evaluate how the same reconstruction would feel when equipped with a variable ratio rudder travel limiter system that maintains full rudder pedal travel at all airspeeds. During VMS runs in which a variable ratio limiter system was simulated, some participants felt that the movements of the pedals were so fast that it was hard to keep their feet on the pedals as they moved.

The group concluded that the VMS, while constrained by certain inherent limitations, provided insight and was a beneficial tool for experiencing time synchronized motions, flight control motions, and displays as opposed to just looking at tabular or charted data. The VMS proved to be an important tool for observing the perceptual cues experienced by the pilots during the accident sequence and assessing the appropriateness of the pilot's inputs and the possible contribution of the rudder system design. These all proved to be central issues in the investigation.

Another central issue in the investigation was the potential role of simulator training in large aircraft upset recovery taught to the pilots. To assess this training, the human performance group conducted a study in the American Airlines A310/300 training simulator to examine the excessive bank angle recovery exercise that the accident pilots completed (NTSB, 2004). Six pilots from the group performed the exercise six times, employing different recovery techniques each time. In the first case, the simulator instructor set up the exercise to replicate the simulator training that pilots received before this accident occurred. The pilots were told they were departing behind a 747 and the instructor initiated an upset when the airplane was banked at an altitude between 2,000 and 2,500 feet and traveling about 240 knots. During the upset, the simulator underwent an uncommanded roll that was randomly set to be either to the left or the right, followed immediately by a large uncommanded roll in the opposite direction. The simulator momentarily inhibited the airplane's response to pilot wheel and rudder pedal inputs during the event to allow the airplane to reach a substantial bank angle before recovery began. Pilots were instructed to recover the airplane according to the method described in the American Airlines advanced aircraft maneuvering program (AAMP), using simultaneous, coordinated rudder in conjunction with control wheel inputs. Each pilot repeated the procedure five additional times, except the roll maneuver was initiated during level flight after the pilot indicated his readiness. The pilots were instructed to use each of the following five recovery methods: partial wheel and no rudder, full wheel and no rudder, full wheel and partial rudder, full wheel and full rudder, and the pilot's preference.

In the AAMP recovery method trials, all of the pilots responded with a full control wheel input (between 77° and 80°) supported by a rudder pedal input (ranging from 6.7° to 14.5° with an average of 10.8°). Five of the six pilots used the rudder pedal simultaneously with the control wheel. Three of the pilots recovered before the airplane reached a 90° bank angle, and the other three pilots recovered the airplane with a maximum bank angle between 108° and 114°. Four of the pilots stated that they were surprised by the onset of the event. The four other prescribed recovery methods showed little difference in average maximum bank angle reached before recovery (between 104° and 107°), and none of pilots recovered before the airplane reached a bank angle of 100°. Three of the six pilots reported that partial wheel and no rudder was the worst recovery method, and all six pilots questioned whether this method provided sufficient control authority for recovery. Two of the pilots felt that a recovery with full wheel and full rudder was the worst method because it created a potential to overcontrol. Data from the full wheel and full rudder recovery suggested a discrepancy between the simulator and the airplane concerning compliance in the rudder control system. Specifically, at 240 knots, the maximum pedal travel on the A300-600 should be limited to 7.9°. When the pilots made full rudder inputs, the maximum pedal travel varied from 10.3° to 18.9°. Some of the pilots reported that they were not able to perceive pushing past the pedal stop when making full pedal inputs in this condition. When the pilots were allowed to recover using their own technique, most of the pilots responded with nearly full wheel and partial rudder pedal inputs. Slightly less input was made on both controls compared to trials where pilots were told to use the AAMP recovery technique, and the pedal response was typically delayed by at least 1 second from the initial control wheel input. The pilots demonstrated a preferred recovery strategy of full wheel and limited rudder in response to the simulator exercise. Also, five of the six pilots indicated, at least once during the six trials, that there was a lack of flight control response during the initial upset.

As with the Bombardier Challenger test flight accident, flight simulation proved to be an invaluable tool during the investigation of the crash of American Airlines Flight 587. In this high-profile accident, unique facilities were used and specific simulator studies were tailored to address specific human performance issues that were ultimately determined to be causal and contributing factors in the crash.

Considerations for Using Flight Simulation as an Investigative Tool

Building on these experiences, agencies responsible for accident investigations, like the NTSB, may consider drafting guidelines for evaluating and conducting flight simulator activities during the course of aviation accident investigations.

While simulation is a valuable tool that will likely have increasing importance in future accident investigations, there are some important considerations to bear in mind that could be formalized through specific guidelines. First, investigators should realize that conducting a simulation is costly and resource intensive. Therefore, the benefits to be derived from simulation should be weighed against these costs and, when appropriate, less costly alternatives considered. Also, in designing a simulation, human performance investigators should understand the limitations of simulation in general and the comparative capabilities, advantages, and disadvantages of different simulators. For example, training simulators may be beneficial in examining how accident flight crews were trained, or examining procedural issues in an accident, but are not well suited for examining highly dynamic motions and accelerations. If the acceleration forces experienced during an accident event are of particular interest, then a unique research facility like the VMS may be needed. Also, if flight control issues arise, training simulators may be inappropriate because they typically lack the fidelity and detail to accurately portray the aircraft's aerodynamic models to the extent that can be achieved in an engineering simulator. In sum, all simulators have limitations and these limitations should be carefully considered when deciding if flight simulation is appropriate and evaluating which particular simulator is best suited to meet the needs of an investigation. Specific simulator limitations should also be identified and considered when designing and executing a simulation plan so that member of the investigative team can understand how these limitations may affect their experiences and the conclusions that can be drawn from the simulation study.

Human performance investigators should also bear in mind some considerations that may limit their ability to answer fundamental questions regarding accident causation in the simulator. First, finding naïve participants may be extremely difficult if not impossible following a high-profile accident. For example, in the American Airlines Flight 587 accident, media coverage and NTSB recommendations focused attention on pilot use of the rudder pedal making it impractical to carry out a simulator study evaluating how certain populations of pilots might use rudder in response to aircraft upsets. Another thing to bear in mind is that some broader issues that arise may be beyond the scope of a focused accident investigation. For example, the American Airlines Flight 587 accident raised many interesting questions regarding the interaction of pilot rudder inputs and rudder system design that would require a large-scale research study to fully address. Formal guidelines for using flight simulation could help investigators better define the purpose, scope, and limitations of simulator activities conducted as part of an aviation accident investigation.

Disclaimer

This paper is based on the author's activities as a Senior Human Performance Investigator for the NTSB. The views expressed in this report are the author's and do not necessarily reflect those of the NTSB or the Congressional Research Service.

References

NTSB (2004, April 14). *Aircraft Accident Brief, CHI01MA006.* NTSB/AAB-04/01.

NTSB (2002, May 15). *Human Performance Study Report*. NTSB Docket Number CHI01MA0006.

NTSB (2002, October 3). *Human Performance Study Report, Vertical Motion Simulator Activities Phase 1: Backdrive of Accident Flight.* NTSB Docket Number DCA02MA001,

NTSB (2004). *Aircraft Accident Report: In-Flight Separation of Vertical Stabilizer, American Airlines Flight 587, Airbus Industrie A300-605R, N14053, Belle Harbor, New York, November 12, 2001*. NTSB/AAR-04/04.

NASA Ames Research Center (2000, October). *VMS Overview: Vertical Motion Simulator Research Facility*. Moffett Field, CA.

Cooper, G. E. & Harper, R. P., Jr. (1969). *The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities*. NASA TN D-5153, 1969.

TOWARDS WIDE-FIELD DISPLAY OF THE GRIPEN HUD INTERFACE TO COMBAT SPATIAL DISORIENTATION

Lars Eriksson and Katarina Undén
FOI, Swedish Defence Research Agency, Dept. of MSI
Linköping, Sweden
Claes von Hofsten
Uppsala University, Dept. of Psychology
Uppsala, Sweden

The head-up display (HUD) interface of the Gripen fighter aircraft utilizes a sphere concept for supporting attitude awareness or spatial orientation (SO). With the sphere interface fixed to the gravitational vertical and the attitude variant aircraft positioned in the center of the sphere, the HUD field-of-regard scans parts of the sphere inside. The HUD interface depicts segments of latitude circles with meridian markings that convey integrated information of pitch, roll, and yaw. To enhance pilot-in-the-loop maneuvering and SO we suggest a wide field-of-view interface design of the Gripen concept, emphasizing the inclusion of peripheral vision. The suggested interface is subsequently integrated with peripheral visual flow to improve SO primarily in instrument meteorological conditions. Implemented in future head-up flight displays systems it could perhaps contribute to a more successful combating of pilot spatial disorientation.

Introduction

To combat pilot spatial disorientation (SD) in fighter aircraft more effectively is a challenge requiring several types of interventions (e.g. Previc & Ercoline, 2004; Small, Wickens, Oster, Keller, & French, 2004). An evolution towards intuitive and more integrated interfaces is one prerequisite for promoting more reliable and safer pilot peak performance. Interface approaches utilizing several sensory channels play key roles in this respect. Integrated auditory, tactile, and visual displays could have a decisive impact on situation awareness (SA), performance, and perceived spatial orientation (SO) (Bles, 2004; Parker, Smith, Stephan, Martin, & McAnally, 2004; Small et al., 2004; van Erp, Veltman, van Veen, & Oving, 2002; Veltman, Oving, & Bronkhorst, 2004). On the other hand, automatic systems for ground and air collision avoidance (GCAS and ACAS) prevent SD accidents by overriding pilot-in-the-loop control. Peak performance in fighter aircraft nevertheless requires a proactive maneuvering by a pilot in the loop. Thus, these reactive automatic systems do not neutralize the need to enhance the pilot's SA, nor the more specific aim for better support of SO or attitude awareness. Furthermore, the crucial sensory information of external frame of reference and events is visual, and the efforts to improve visual interfaces per se thus continue because of the critical role vision plays.

The risk for SD increases in instrument meteorological conditions (IMC) (e.g. Previc, 2004), and judging pitch and bank by referring to the artificial instruments in fog or darkness is less accurate and compelling than viewing the outside ground with horizon in good visibility (Ercoline, DeVilbiss, & Evans, 2004; Gillingham & Previc, 1993). Thus, the flight instruments or visual interfaces show less than acceptable effectiveness. It can be argued that they ought to be in better resonance with the natural mode of perceiving SO (e.g. Eriksson & von Hofsten, 2005; Leibowitz, 1988; Malcolm, 1984). The interfaces need to intuitively convey integrated information for maneuvering and to generate an accurate and compelling perception of SO (Ercoline et al., 2004; Eriksson, 2005). Along these lines, and anticipating further advances in visual displays technology, we present some ideas aiming for improving pilot-in-the-loop maneuvering and SO. First, we present the basic principles for the head-up display (HUD) interface of the Gripen fourth generation fighter aircraft that conveys integrated information of pitch, roll, and yaw. Second, we apply the Gripen HUD interface to a wide field-of-view (FOV) display format to incorporate peripheral vision. Third, we integrate the interface with flight-adapted peripheral visual flow.

The Gripen HUD interface

Figure 1 illustrates the Gripen HUD interface as principally appearing during horizontal flight in visual meteorological conditions (VMC), including the flight parameters altitude, speed, flight path marker/velocity vector, G-load, angle of attack (AoA), and heading. Horizon-line and "pitch lines" with "yaw markings" are also indicated. (Note: All illustrations of the Gripen HUD interface depict basic principles/configurations and not actual symbology in detail.) The HUD interface incorporates a sphere concept as reference frame for maneuvering and

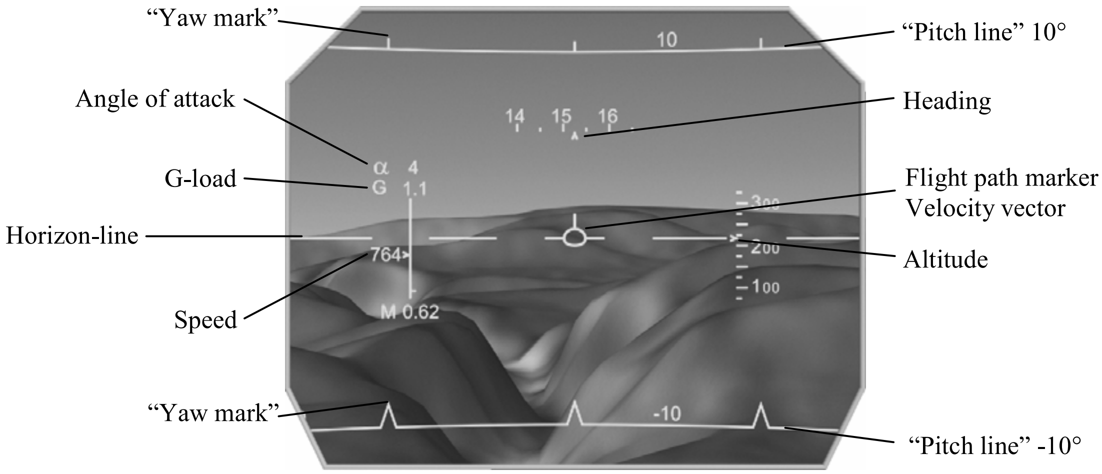


Figure 1. An illustration emphasizing basic principles of the Gripen HUD interface with flight parameters superimposed on the environment in VMC.

perceiving SO with the flight parameters. Although the attitude of the aircraft varies, it is permanently positioned in the center of a sphere that has its vertical axis fixed to the gravitational vertical. The sphere consists of latitude circles at each 10° pitch deviation from the horizontal up to 80°, with the latitudinal great circle of the sphere equal to the horizontal and depicted as a straight line. See horizon-line and "pitch lines" in Figure 1. That is, the HUD depicts segments of latitude circles showing increasing curvature with increasing deviation from horizontal. The full zenith circle is shown when the aircraft is pointing straight up and the nadir circle pointing straight down. Together with meridian markings on the latitude circles, integrated information of pitch, roll, and yaw is conveyed. The meridian markings are different on dive-circles compared to climb-circles to make them easily distinguishable. They could be called "yaw markings" because they indicate yaw position or, more important, change in yaw position. Figure 2 illustrates a pitch-up sequence with no change in yaw or roll position. The sequence goes from horizontal flight with an actual 4° pitch attitude of aircraft, with velocity vector (flight path marker) at 0° and AoA of 4°, to 75° pitch attitude, with velocity vector at 90° and AoA of 15°. Metaphorically put, the sphere concept corresponds to viewing parts of a large ADI ball from its inside (ADI - Attitude Director Indicator).

Operative for quite awhile in the Gripen aircraft, the overall intuitive design and the consistent dynamics of integrated pitch, roll, and yaw have received appreciation from pilots. One aspect of the consistent dynamics is revealed in transitioning from flying upwards in upright orientation to flying downwards

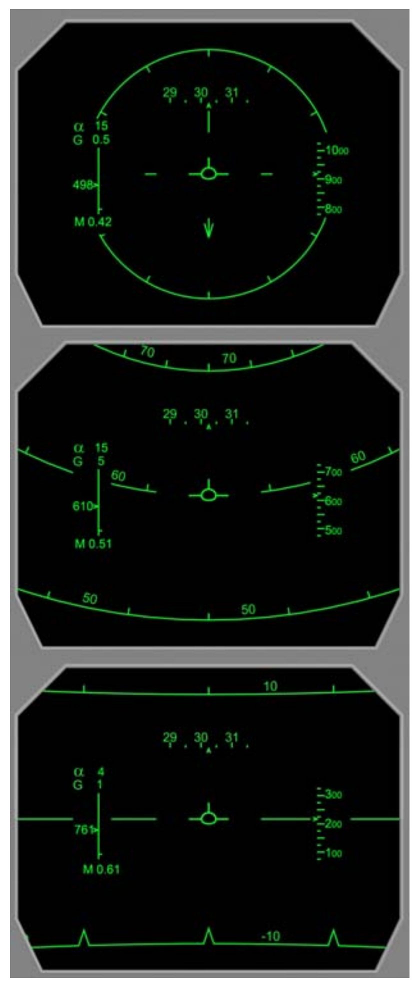


Figure 2. From bottom to top: Pitch attitude of 4° with velocity vector at 0°, 75° pitch-up with velocity vector at 60°, and 75° pitch-up with velocity vector at 90°.

upside-down when performing a looping. In comparison to a "regular pitch-ladder design" that will turn over the up – down orientation of the horizon-parallel line-segments, the sphere interface shows stability. The transition from flying upwards to downwards only means that the HUD field-of-regard transitions smoothly and stable to scanning the opposite side of the sphere, and flying inverted still entails that climb-circles segments bend upwards and dive-circles segments downwards.

A wide FOV interface design

Visual field coverage is of course an important factor in displays developments, and an increased FOV incorporating the peripheral visual field could improve the support of SO (Leibowitz, 1988; Wickens & Hollands, 2000). Because of the constraints for flight displays technology, however, suggestions of interface designs naturally emphasize central vision (e.g. Flach, 1999; Previc & Ercoline, 1999). The Malcolm Horizon projected on the instrument panel and the Background Attitude Indicator (BAI) on head-down displays can be considered exceptions (Comstock, Jones, & Pope, 2003; Ligget, Reising & Hartsock, 1999; Malcolm, 1984). Still, HUDs and helmet mounted displays (HMDs) allowing peripheral visual field presentation to great extent are yet to be realized. However, it is relevant to investigate the fundamentals for an interface design applied to a large FOV display format simply because of the advancement of displays technology.

One disadvantage with the emphasis of current flight displays on central vision is that they therefore primarily depend on directed attention. Furthermore, the functional dichotomization of vision into focal and ambient subsystems represents two separate perceptual modes (e.g. Leibowitz, 1988). The focal processes the most central part of the visual field and the ambient utilizes the entire visual field. Focal is primarily associated with object and event detection/identification and ambient with spatial awareness and SO (linked to the parallel parvo- and magnocellular channels). Information for SO is thus primarily provided by ambient vision that is typically not contingent on attention, and increasing the FOV to include peripheral vision could improve spatial awareness, SO, and the support of maneuvering. In particular, compared to a Malcolm Horizon, or a head-down BAI, a wide FOV utilizing the Gripen HUD concept has the advantage of integrating not only pitch and roll, but also yaw. An illustration of an application of the Gripen concept to a wide FOV is shown in Figure 3.

Goals in the US Air Force Displays Vision include a definition of a "panoramic" class of SA (Tulis, Hopper, Morton, & Shashidhar, 2001). The *"basis for identifying the panoramic SA goal comprises such factors as the excitation of peripheral vision cues for horizontal viewing fields greater than about 100 degrees and the opportunity to present integrated display formats"* (Tulis et al. 2001, p. 11). The Panoramic Night Vision Goggle (PNVG) has accordingly a FOV of about 100° by 40° (horizontal by vertical) (e.g. Geiselman & Craig, 1999; Jackson & Craig, 1999). Interestingly, a PNVG with superimposed computer-generated symbology is also an emergent further development, i.e. symbology overlay on the PNVG mediated night scene made possible by miniature flat panel displays (or similarly). Thus, applications of interface designs extending far outside the central visual field could perhaps include PNVGs, if not HUDs or HMDs.

Peripheral visual flow integration

The risk for SD accidents increases in IMC despite intense training, experience, and hammered-in instructions to fly by the instruments. It seems as if the pilot's perceptual processing is not in contact with crucial factors that contribute to overcoming erroneous perceptions of SO. Display interfaces not only ought to go beyond central visual field in IMC, they ought to utilize the ambient system more effectively. The ambient visual system is primarily in resonance with motion elements grouped over larger areas, as with locomotion generated optic flow (e.g. Gibson, 1966; Johansson & Börjesson, 1989; Lee, 1980). Visual flow (optic flow) can even dominate proprioceptive and equilibrium sense information (e.g. Lishman & Lee, 1973). In particular, flight-adapted visual flow with combined expanding and rotational motions seems to sensitize the visually guided SO system, demonstrating an effective suppression of vestibular and proprioceptive information. (Unless desired to lose balance, one ought to hold onto something standing in a dome fixed platform flight-simulator and viewing a flight maneuver visually represented as a "roll movement of the ground".) A wide FOV interface could utilize an artificial visual flow to suppress erroneously perceived SO based on proprioception and the vestibular sense.

The opto-kinetic cervical reflex (OKCR) involves a lateral tilt of the pilot's head towards the horizon during aircraft roll maneuvers and reveals itself in VMC but not IMC (e.g. Patterson, Cacioppo, Gallimore, Hinman, & Nalepka, 1997). While the spatial frame of reference lies outside the aircraft in

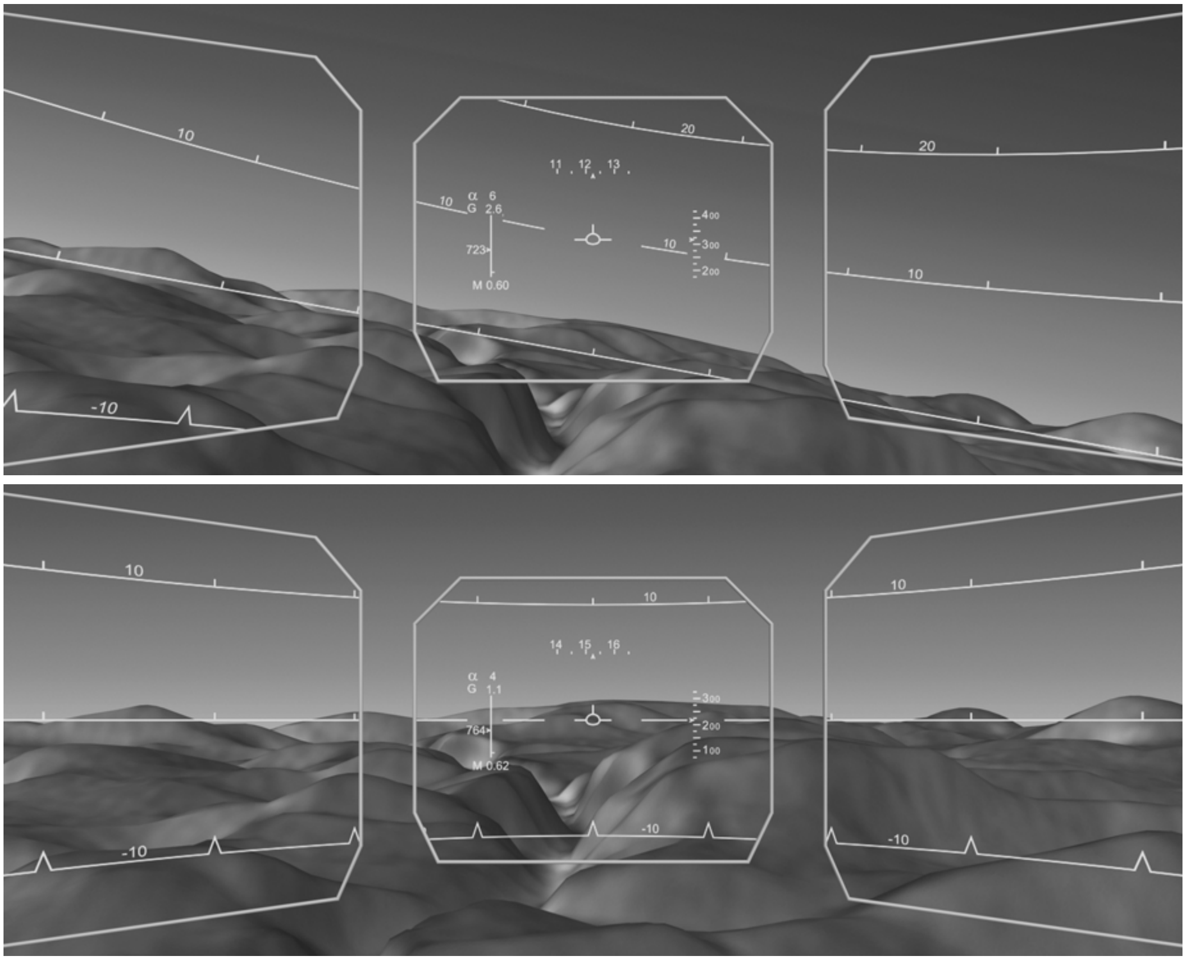


Figure 3. The Gripen HUD concept applied to one version of a wide FOV interface design. The illustrations show horizontal flight at the bottom, and a roll position in a pitch attitude above horizontal at the top.

VMC, it is situated inside the aircraft in IMC (Johnson & Roscoe, 1972; Patterson et al., 1997). A presentation of an artificial peripheral visual flow combined with conformal horizon-line information in central visual field could perhaps trigger the same sensory reflexes in IMC as occur in VMC (e.g. Eriksson, 2005; Eriksson & von Hofsten, 2003, 2005). Figure 4 illustrates a flight sequence with an IMC mode of the suggested interface that includes flight-adapted peripheral visual flow, i.e. visual flow represented by the black & white textured ground. Improved spatial awareness and lowered mental workload could be some of the effects of a triggered OKCR in IMC (see Patterson et al., 1997, for a qualitative model of SO in VMC and discussion of HMD design). On the other hand, a pilot must *"refer to the instrument displays in both good and bad weather conditions in order to fly the aircraft safely"* (Ercoline et al. 2004, p. 382) in that air speed and altitude, for example, are particularly difficult to extract from perceiving the outside world or an artificial visual flow. This is most important during low-level flight to avoid controlled-flight into terrain. The peripheral visual flow integrated interface includes these parameters by utilizing the Gripen avionics system (Figure 1). Furthermore, while the ground proximity warning complements an automatic GCAS, the rate of the auditory stall warning enhances the pilot's proactive performance by indicating the stall margin (cf. Flach, 1999).

Concluding remarks

The utilization of an operative HUD interface concept integrating information of pitch, roll, and yaw provides the important fundamentals of an integrated reference frame for maneuvering. The suggested wide FOV interface design seems to have two advantages for further enhancing pilot-in-the-loop maneuvering and SO. First, the wide FOV

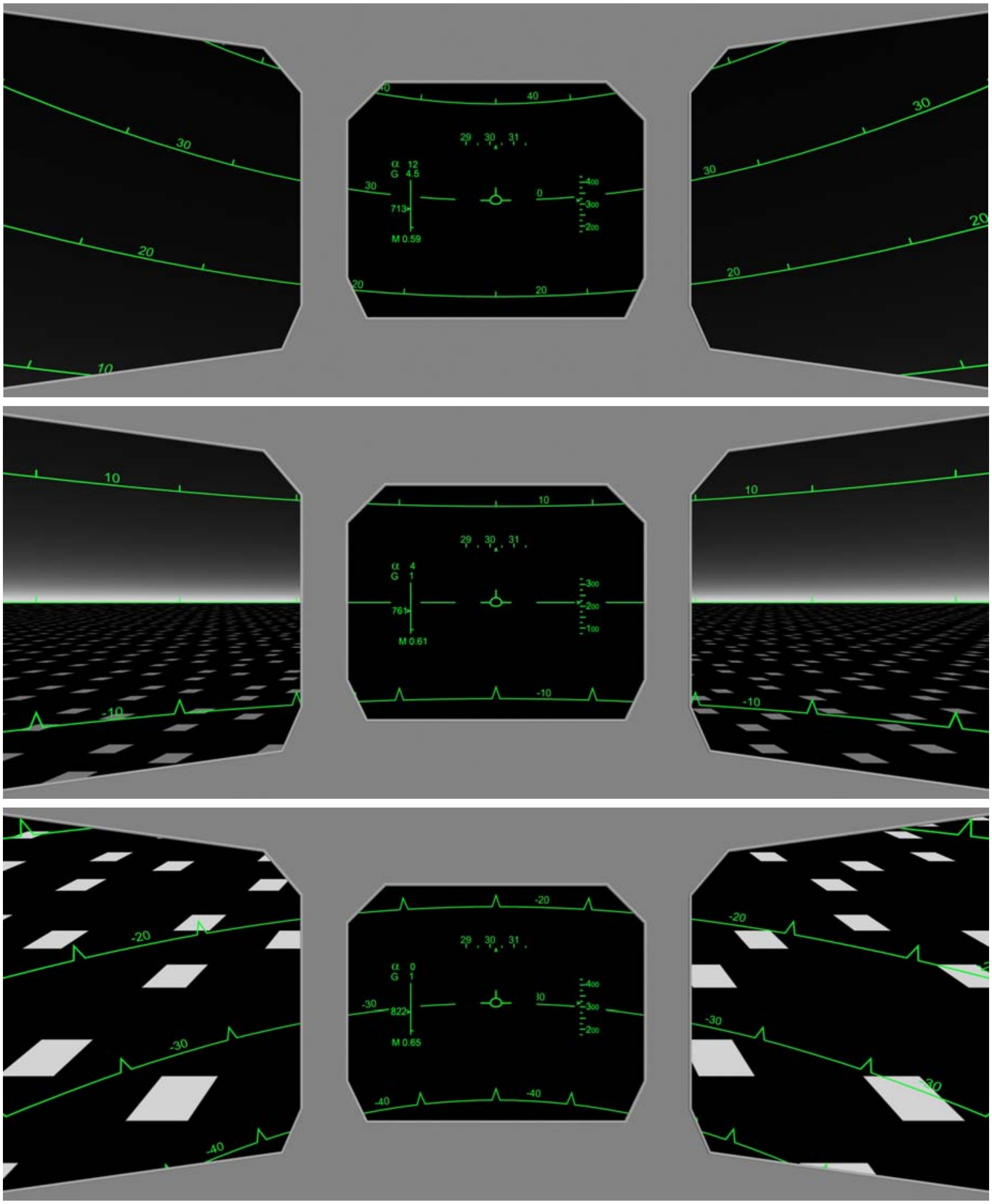


Figure 4. An illustration of an IMC mode of the suggested interface with peripheral visual flow. The flight sequence from bottom to top: From pitch-down to pitch-up including horizontal flight in between.

inclusion of peripheral vision supports perception of SO and maneuvering more effectively. Second, peripheral visual flow is integrated into the sphere concept in a geometrically correct configuration, enhancing visual resonance with the SO mechanism primarily in IMC. Accordingly, it seems to show potential for triggering sensory reflexes critical for SO, reinforcing information for maneuvering, and capturing pilot attention when transitioning into critical aircraft attitudes. It could therefore contribute to a more effective combating of pilot SD in the future.

The ideas presented here emphasize basic concepts that of course need refinement. Head-up flight displays systems allowing a wide FOV interface design are also yet to be realized. On the other hand, the design can be implemented and subjected to empirical scrutiny by experiments carried out in research applications platforms. Another issue is that the Gripen HUD interface provides an intuitive visual frame of reference for three-dimensional cueing with auditory and tactile displays, supporting multisensory approaches to improve pilot peak performance.

Acknowledgments

This research was supported by the Swedish Armed Forces. Jimmy Berggren is gratefully acknowledged for producing the illustrations, significantly contributing to the aim for making ideas (appear to) fly. We also thank Örjan Skinnars, Saab AB, for his always-friendly support, Anders Persson, Saab AB, Jens Alfredson, FOI, and Staffan Nählinder, FOI.

References

Bles, W. (2004). Spatial disorientation countermeasures – advanced problems and concepts. In F.H. Previc & W.R. Ercoline (Eds.), *Spatial disorientation in aviation* (pp. 509-540). Reston, VA: American Institute of Astronautics and Aeronautics, Inc.

Comstock, J.R., Jones, L.C., & Pope, A.T. (2003). The effectiveness of various attitude indicator display sizes and extended horizon lines on attitude maintenance in a part-task simulation. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, (pp. 144-148), 13-17 October, Denver, USA.

Ercoline, W.R., DeVilbiss, C.A., & Evans, R.H. (2004). Flight displays I: Head-down display topics for spatial orientation. In F.H. Previc & W.R. Ercoline (Eds.), *Spatial disorientation in aviation* (pp. 379-449). Reston, VA: American Institute of Astronautics and Aeronautics, Inc.

Eriksson, L., & von Hofsten, C. (2003). Effects of peripheral visual flow on postural responses and implications for a spatial orientation visual interface in aircraft. In *Proceedings of the 12th International Conference on Aviation Psychology*, 14-17 April, Dayton, OH, USA.

Eriksson, L., & von Hofsten, C. (2005). Effects of visual flow display of flight maneuvers on perceived spatial orientation. *Human Factors*. (In press.)

Eriksson, L. (2005). Towards a peripheral visual flow integrated display format to combat pilot spatial disorientation. (Manuscript in preparation.)

Flach, J.M. (1999). Ready, fire, aim: A “meaning-processing” approach to display design. In D. Gopher & A. Koriat (Eds.), *Attention and Performance XVII* (pp. 197-222). Cambridge: MIT Press.

Geiselman, E.E., & Craig, J.L. (1999). Panoramic night vision goggle update. In *Proceedings of the 1999 SAFE Symposium*, Atlanta, GA, USA.

Gibson, J.J. (1966). *The senses considered as perceptual systems.* Boston, MA: Houghton Mifflin.

Gillingham, K.K., & Previc, F.H. (1993). *Spatial orientation in flight*. Technical Report AL-TR-1993-0022, Armstrong Laboratory, Crew Systems Directorate, Crew Technology Division, Brooks Air Force Base, TX, USA.

Jackson, T.W., & Craig, J.L. (1999). Design, development, fabrication, and safety-of-flight testing of a panoramic night vision goggle. In *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE): Head-Mounted Displays IV, 3689*, (pp. 98-109), SPIE 3689.

Johansson, G., & Börjesson, E. (1989). Toward a new theory of vision: Studies in wide-angle space perception. *Ecological Psychology, 1(4)*, 301-331.

Johnson, S.L., & Roscoe, S.N. (1972). What moves, the airplane or the world? *Human Factors, 14(2)*, 107-129.

Lee, D.N. (1980). The optic flow field: The foundation of vision. *Philosophical Transactions of the Royal Society of London, B, 290*, 169-179.

Leibowitz, H.W. (1988). The human senses in flight. In E.L. Wiener & D.C. Nagel (Eds.), *Human Factors in Aviation* (pp. 83-109). San Diego: Academic Press Inc.

Liggett, K.K., Reising, J.M., & Hartsock, D.C. (1999). Development and evaluation of a background attitude indicator. *The International Journal of Aviation Psychology, 9(1)*, 49-71.

Lishman, J.R., & Lee, D.N. (1973). The autonomy of visual kinaesthesis. *Perception, 2*, 287-294.

Malcolm, R. (1984). Pilot disorientation and the use of a peripheral vision display. *Aviation, Space, and Environmental Medicine, 55*, 231-238.

Parker, S.P.A., Smith, S.E., Stephan, K.L., Martin, R.L., & McAnally, K.I. (2004). Effects of supplementing head-down displays with 3D-audio during visual target acquisition. *International Journal of Aviation Psychology, 14(3)*, pp. 277-295.

Patterson, F.P., Cacioppo, A.J., Gallimore, J.J., Hinman, G.E., & Nalepka, J.P. (1997). Aviation spatial orientation in relationship to head position and attitude interpretation. *Aviation, Space, and Environmental Medicine, 68(6)*, 463-471.

Previc, F.H. (2004). Visual illusions in flight. In F.H. Previc & W.R. Ercoline (Eds.), *Spatial disorientation in aviation* (pp. 283-321). Reston, VA: American Institute of Astronautics and Aeronautics, Inc.

Previc, F.H., & Ercoline, W.R. (2004). *Spatial disorientation in aviation.* Reston, VA: American Institute of Astronautics and Aeronautics, Inc.

Previc, F.H., & Ercoline, W.R. (1999). The “outside-in” attitude display concept revisited. *The International Journal of Aviation Psychology, 9(4)*, 377-401.

Small, R.L., Wickens, C.D., Oster, A.M., Keller, J., & French, J.W. (2004). *Multisensory integration for pilot spatial orientation.* Report AF03-061, AFRL/HECI, Crew Systems Development Branch, Wright Patterson AFB, OH, USA.

Tulis, R.W., Hopper, D.G., Morton, D.C., & Shashidhar. (2001). Review of defense display research programs. In *Proceedings of SPIE, Cockpit Displays VIII: Displays for Defense Applications, Vol. 4362*, pp. 1-25, 2001.

van Erp, J.B.F., Veltman, J.A., van Veen, H.A.H.A.C., & Oving, A.B. (2002). Tactile torso display as countermeasure to reduce night vision goggles induced drift. In *Proceedings of NATO RTO HFM Panel Symposium on Spatial Disorientation in Military Vehicles: Causes, Consequences and Cures*, 15-17 April, La Coruña, Spain.

Veltman, J.A., Oving, A.B., & Bronkhorst, A.W. (2004). 3-D audio in the fighter cockpit improves task performance. *International Journal of Aviation Psychology, 14(3)*, pp. 239-256.

Wickens, C.D., & Hollands, J.G. (2000). *Engineering psychology and human performance.* Third ed. Upper Saddle River, NJ: Prentice Hall.

EVALUATING THE ACCIDENT RATE: CULTURAL AND ORGANIZATIONAL FACTORS AS MAJOR CHALLENGES IN THE IMPROVEMENT OF AVIATION SAFETY–ANALYSIS OF RELATED ACCIDENTS

Eugenio L. Facci
Department of Psychology
University of Central Florida

Meredith A. Bell
Department of Psychology
University of Central Florida

Razia V. Nayeem
Department of Psychology
University of Central Florida

This paper aims at examining the accident rate in the aviation industry and specifically at understanding why the accident rate has been decreasing at a gradually slower rate in the past few decades. We highlight an emerging trend of accidents where major safety threats were allowed to remain unnoticed and unaddressed until a crash eventually occurred, with the further goal of analyzing the apparent difficulty that the industry has in effectively handling and addressing safety threats before crashes occur. In the article we analyze the organizational factors that most likely allow this detrimental phenomenon to take place and we examine what guidelines can be found in the Organizational Behavior literature to help examine and develop potential countermeasures.

Introduction

If you were to decide whether 160 lbs is an accurate estimate of the average US airline passenger weight, would you think that this value is definitely correct or would you think that it may be safer to verify this value? We believe that most people, and definitely most pilots, would want to verify the value and make sure that they are not flying a heavy or unbalanced aircraft. However, after the crash on takeoff of an Air Midwest Beech 1900 in Charlotte, NC in January 2003, investigators came across this startling finding. The carrier estimated that the average US airline passenger weight was 160 lbs, and the value of 160 lbs was the one used in their daily flight planning operations[1]. The actual average US airline passenger weight was found to be, in a consequent survey conducted by the FAA, 186 lbs[2], more than 20 lbs in excess of the erroneous estimation that had been in use for years. This finding, along with similar ones from other recent airline accident investigations, brings forward the question of why large errors and misconceptions are allowed to exist for prolonged times in the aviation industry. In this paper we argue that accidents of this type, which highlight inadequate, confusing or inexistent guidelines and, in general, poor industry-wise communication processes, are a growing problem for the aviation industry. We also maintain that this arising trend may help explain why accident records are not significantly improving any more, and that a more proactive safety approach may be needed in order to cope with the problem.

The paper is structured as follows. Section one discusses recent airline accidents exemplifying the accident trend that, we believe, is taking hold in the industry. Section two quantifies this trend. In part three and four we review the relevant literature and provide our own view in an attempt to explain why this phenomenon is taking place. Section five examines what are the implications of our historical analysis of accident trends, and in section six we draw conclusions and indicate potential streams of research for future inquiry.

Recent Accidents

In this section we provide three examples of crashes that, we believe, highlight the industry's limited capability to deal with misconceptions and inadequacies before accidents happen.

[1] The value of 160 lbs was used during summer and 165 lbs during winter.
[2] The NTSB accident report quotes a value of 196 lbs, since 10 lbs were added for personal items.

- American Airlines 587 - The morning of November 12th, 2001, flight 587, an Airbus 300, lost the vertical tail fin and the engines shortly after taking off from JFK airport in New York. The NTSB concluded that the First Officer induced with the rudder a series of side oscillations that caused the tail to exceed its design limits and separate. According to the Board the pilot was under the erroneous impression that the plane movements were caused by turbulence and not by his own control inputs, so that he initiated and exacerbated an aggressive action on the flight controls until structural failure occurred. The Board investigation found that the underlying basis for this accident was provided by wide misconceptions diffused among line pilots. Specifically, pilots seemed to believe that they could apply series of full rudder inputs when flying below Maneuvering Speed V_a. In actuality only one full application is allowed when followed by bringing the controls back to neutral. Furthermore, pilots seemed not to be aware that, in the A 300, pedal force and pedal travel required to achieve full rudder deflection decreases as speed increases. Finally, most pilots seemed also to believe that a control limiter system in the plane would protect the frame from excessive aerodynamic loads under all circumstances, which again was an incorrect belief. As soon as these misconceptions were discovered, the NTSB issued prompt warnings to correct them.
- B 727 FedEx crash in Tallahassee, FL, July 2002 – The plane impacted the ground before the approach end of the runway as the first officer flew an excessively low final approach segment. The NTSB found that the first officer suffered from a colorblind vision impairment, which caused him to be unable to use the red and white PAPI lights, the only available aid on the runway to which the approach was being conducted. The First Officer's medical certificate had been issued according to a FAA waiver for colorblind vision requirements. In the accident report, the NTSB states that "It is apparent that in some situations, accurate color vision may be critical to a degree that is not currently reflected in the application of the aviation medical certification standards…The Safety Board notes that current aviation medical certification standards for color vision and related screening tests do not emphasize the full complexity of color in modern operational situations…"
- Air Midwest 5481, Charlotte, NC, January 8th, 2003 – The plane lifted off and kept pitching up until a stall occurred. Loss of control and ground impact ensued. Investigators found that the plane was loaded with an excessively aft center of gravity, partially due to the carrier's assumption that each passenger weigh 165 lbs. A survey conducted soon afterwards revealed that 186 lbs was a more accurate estimation of the average passenger's weight. Investigators came to the conclusions that several other flights had been conducted in an overweight or unbalanced condition and that the incorrect estimations had been in use for years. However, on the accident day the problem was exacerbated by a maintenance error which reduced the pitch-down authority of the elevator control in the accident plane.

Data

In our analysis, we attempted to separate accidents where procedures or guidelines were not followed – i.e. "The captain's failure to abort the approach when stabilized approach criteria were not met" – from accidents where procedures themselves were found to be the source of the problem. We specifically looked for accidents where inadequate, confusing or inexistent procedures and guidelines were cited by the NTSB as a causal or contributing factor – for the purpose of our analysis, we called those events Procedure/Guideline Related Accidents (PGRAs)[3]. Data suggest that this latter type of accident is becoming a growing problem for the industry. In the period 1996-2003, eighty percent of fatal Part. 121 (Air Carrier) accidents were PGRAs, up from 53% for the period 1989-1996 and 43% for 1982-1989. A three-year breakdown analysis yields similar results. (Tables 1,2,3 and 4).

[3] Going back to our initial question "why are large errors and misconceptions allowed to exist for prolonged times in the aviation industry?", it is useful to highlight how inadequate procedures are the manifestation of errors – i.e. wrong W&B estimations – while confusing procedures are the cause of misconceptions.

Literature

The issue of why errors and misconceptions may be allowed to persist for prolonged times within an organization has been studied in literature primarily by looking at what causes employees to be more or less likely to express their safety concerns, and by looking at how responsive are different types of organizations to the emergence of new accident trends. Hofmann and Morgenson (1999) examined the influence of Leader Member Exchange (LMX) on safety communication, where high-quality LMX involved minimal power distance between leaders and subordinates, open discussion of non-routine problems, joint decision making and strong value congruence. They found that high quality LMX was associated with openness to the raising of safety concerns (see also Fairhurst, Rogers & Sarr, 1987; Liden, Sparrowe, & Wayne, 1997; Simard & Marchand, 1997). Along the same lines, Edmonson (1996) indicated that positive safety climates, which he conceived as the result of high importance placed on safety, high management's commitment to safety and non-punitive approaches to accident investigations, will lead to free-flowing information exchange about safety matters (see also DeJoy, 1985). Regarding more specifically the topic of organizational responsiveness Milliken et al. (1998) highlighted that different industries show different degrees of responsiveness to critical issues. De la Cruz Déniz-Déniz & de Saá-Pérez (2003) also came to the interesting finding that organizations that follow social responsibility principles – and therefore place a high level of importance on the individuals and their contribution – typically show a high level of organizational responsiveness. From a comprehensive perspective the literature seems to provide cohesive indications that, in order for employees to openly express their safety concern, there must be a high value placed on safety by the organization (positive safety climate) and there must be an unrestrained communication flow, which the literature sometime defines as a positive Leader Member Exchange. However, other factors may need to be taken into account to explain individual and organizational safety attitudes, especially in the aviation industry, and the goal of the next section of our paper is to discuss these factors.

The Dark Side of Procedures

Based on existing literature and on evidence provided by a number of accident reports, we believe that three main factors may help explain the industry's limited capability to deal with errors and inadequacies on a pre-emptive basis. These factors act both at the individual level and at the organization's level.

Individual Level

- Reluctance – In a setting where the concept of procedure is highly regarded, it may be difficult for employees to make the point that some procedure is inadequate, or confusing. Quite easily, their own competence may be called into question. Or, along the same lines, their action may be perceived by colleagues as unwillingness to comply with existing procedures.
- Social proof – In a large system where all employees are trained in a highly standardized manner, there may be a tendency for any single individual who comes across a potential procedural glitch to think that someone else already noticed and brought up the issue, so that further intervention is not needed.

Organizational level

- Inertia – An organization that strongly relies on the concepts of "procedure" and "guideline" may look for guidelines even when new phenomena arise, such as a new type of accident trend. This may delay the realization that something new is happening and that the adoption of new counteracting strategies is needed.

Implications - A Historical Perspective

Throughout the years, the focus of aviation safety has mostly been on creating knowledge to resolve technical problems. In the 1950s, in-flight explosions of Comet jets, for instance, pointed to the unknown danger imposed by pressurization cycles on aircraft frames. In the 1970s, windshear accidents and Controlled Flight Into Terrain accidents (CFITs) led researchers to look into issues such as microburst activity on one side and crew coordination techniques on the other. However, as the system grew more and more complex, the key issue shifted from creating knowledge that could help address safety-sensitive issues to managing the amount of knowledge that had been created. The recent accidents that we examined in fact highlighted threats that could have probably been solved before the accidents took place. That is, the capabilities needed to analyze and resolve the issues were available before and at the time the accidents occurred. However, nobody spoke up, and the actual failure occurred not at the level of understanding or resolving the problems but at the level of detecting them. To this regard, the pattern of causality of today's accidents and accidents from earlier decades seems to be remarkably different. In the 1970s and 1980s windshear accidents happened repeatedly because we did not know exactly what windshear was or, regarding the issue of CFITs, what were the best crew coordination techniques. Figuring out those problems and their solutions took years of research efforts. In the case of recent accidents, instead, the capabilities needed to address the relevant safety issues were already available, such as the capability to determine a realistic average passenger's weight or the capability to teach pilots how to correctly use rudder. However, that critical piece of information did not make it to the right people, so that the critical failure occurred at the level of diffusing safety-sensitive knowledge within the aviation system rather than at the level of creating that knowledge. The industry's current difficulty in managing safety information, specifically, seems to be bidirectional – how do we get employees to point out safety problems ("Is 160 lbs a correct estimation?") and how do we get information to the relevant employees – ("do not apply series of full rudder inputs in opposite directions even under V_a"). In order to fully address these problems we may need to climb up the chain of events that leads to an accident and consider the "soft", underlying and intangible factors that create the conditions for an accident to take place. For instance, after determining that incorrect Weight and Balance data played a role in the Air Midwest January 2003 crash, we would then need to understand what social dynamics made It possible for the use of incorrect data to go undetected for years. Only then we would be able to go to the root of the problem and we would therefore be able to decrease the likelihood that similar events may occur in the future (Figure 1). Along the same line of reasoning, it should also be noted that the traditional Human Factors and Engineering approach to aviation safety may need to be expanded to a broader view, which may include issues more typically studied in the Organizational Behavior and Management study areas, such as "How do we detect safety threats?" and "How do we get critical information to the relevant employees?" (Figure 2)

Conclusions

In this paper we argue that the aviation industry is faced with an arising trend of accidents for which procedures and guidelines are found to be causal or contributing factors. We maintain that the industry's ability to address those crashes will directly depend on the ability to manage safety-sensitive information, and that in general the industry may benefit from improving the tangible and intangible frameworks on which communication processes rely. Our reasoning is based on an analysis of all fatal US Air Carrier accidents from 1982 to 2003 and on a detailed examination of three among the most recent major NTSB investigations. Indications for future research are also discussed.

Timeframe (7 years)	# Part. 121 Accidents	# PGRAs	% PGRA/Accidents
Jan. 1996-Jan. 2003	10	8	80%
Jan. 1989-Jan. 1996	19	10	53%
Jan. 1982-Jan. 1989	21	9	43%

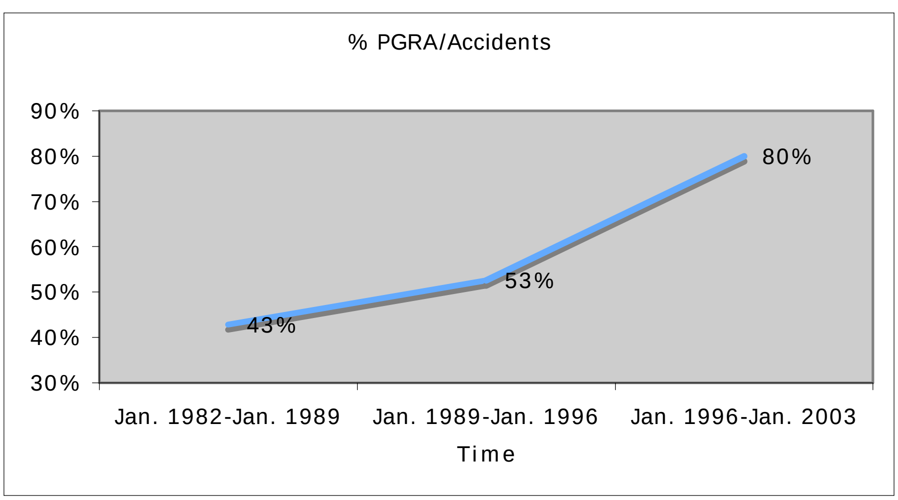


Table 1 & 2. *Number of PGRA Accidents in Part. 121 operations from 1982 to 2003, seven-year breakdown.*

Timeframe (3 years)	# Part. 121 Accidents	# PGRAs	% PGRA/Accidents
Jan. 2000-Jan. 2003	4	3	75%
Jan. 1997-Jan. 2000	2	2	100%
Jan. 1994-Jan. 1997	8	5	63%
Jan. 1991-Jan. 1994	5	2	40%
Jan. 1988-Jan. 1991	12	8	67%
Jan. 1985-Jan. 1988	10	5	50%
Jan. 1982-Jan. 1985	9	2	22%

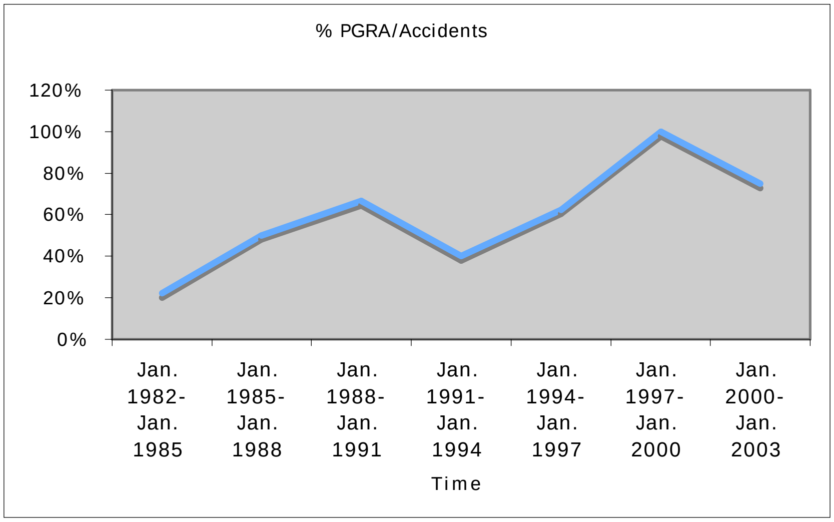


Table 3 & 4. *Number of PGRA Accidents in Part. 121 operations from 1982 to 2003, three-year breakdown.*

Figure 1. *Climbing up the causal chain*

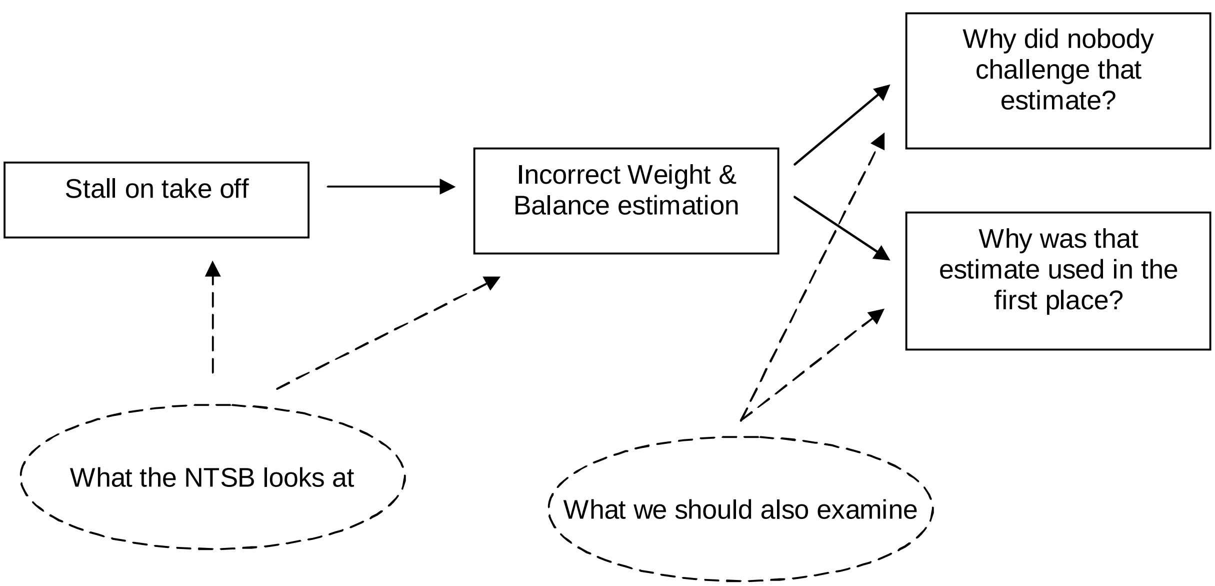
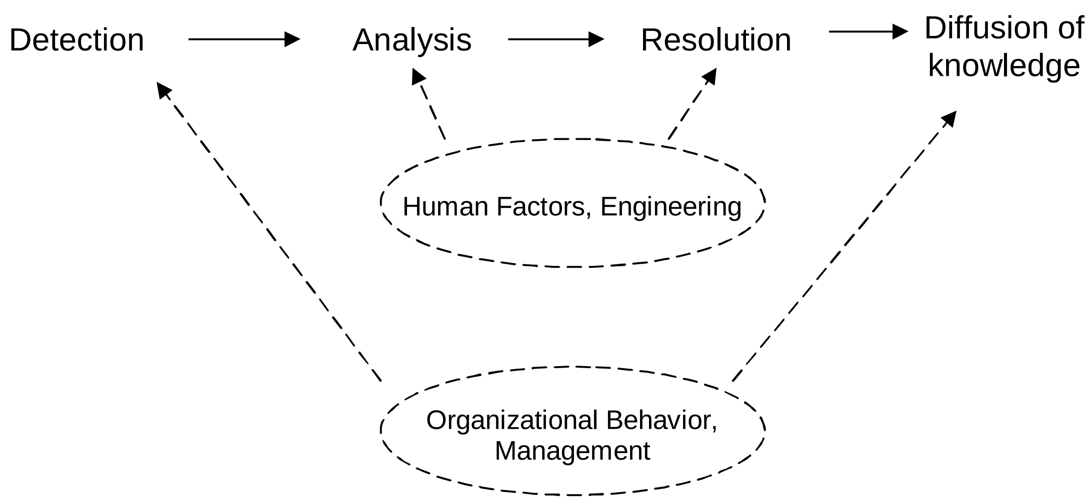


Figure 2. *Managing safety threats*



References

Cialdini, R. (1993). *Influence: The Psychology of Persuasion.* New York: Quill, William Morrow

Columbia Accident Investigation Board (2003). *Report Volume 1.* Washington, DC. US Government Printing Office

De la Cruz Déniz-Déniz, M. & de Saá-Pérez, M. (2003). A Resource-based View of Corporate Responsiveness toward Employees. *Organization Studies, 24(2),* 299-319.

DeJoy, D. M. (1985). Attributional processes and hazard control management in industry. *Journal of Safety Research, 16,* 61-71.

Edmonson, A. C. (1996). Learning from mistakes is easier said than done: Group and organizational influences on the detection and correction of human error. *Journal of Applied Behavioral Science, 32,* 5-28.

Facci, E. L., Sumwalt, R., Bell, M. A. (2005). *Creating and sustaining a positive safety climate. From reaction to proaction – ways to improve the status quo.* Proceedings of the 17th European Aviation Safety Seminar, Warsaw, Poland, March 2005. Alexandria, VA: The Flight Safety Foundation

Fairhurst, J. T., Rogers, L. E., & Sarr, R. A. (1987). Manager-subordinate control patterns and judgments about the relationship. *Communication Yearbook, 10,* 395-415.

Hofmann, D. A., & Morgenson, F. P. (1999). Safety-related behavior as a social exchange: The role of perceived organizational support and leader-member exchange. *Journal of Applied Psychology, 84,* 286-296.

Liden, R. C., Sparrowe, R. T., & Wayne, S. J. (1997). Leader-member exchange theory: The past and potential for future. *Research and Personnel and Human Resource Management, 15,* 47-119.

Milliken. F. J., Martins, L. L., Morgan, H. (1998). Explaining organizational responsiveness to work-family issues: The role of human resource executives as issue interpreters. *Academy of Management Journal, 41(5),* 580-592.

Morrison, E. W. (2000) *Organizational silence: A barrier to change and development in a pluralistic world.* Academy of Management Review, October 1, 2000, Vol. 25, Issue 4

National Transportation Safety Board (2004). *American Airlines Flight 587 Belle Harbor, New York November 12, 2001.* The National Transportation Safety Board, Washington DC.

National Transportation Safety Board (2004). *Federal Express Flight 1478 Tallahassee, Florida July 26, 2002.* The National Transportation Safety Board, Washington DC.

National Transportation Safety Board (2003). *Air Midwest Flight 5481 Charlotte, North Carolina January 8, 2003.* The National Transportation Safety Board, Washington DC.

Simard, M., & Marchand, A. (1997). Workgroups' propensity to comply with safety rules: The influence of macro-organizational factors. *Ergonomics, 40,* 172-188.

DECISION MAKING DURING AN AIRLINE RESCHEDULING TASK: A CONTEXTUAL CONTROL MODEL DESCRIPTION

Karen M. Feigh, **Amy R. Pritchett**, **Julie A. Jacko**, **Tina Denq**
Georgia Institute of Technology
Atlanta, Georgia

This paper examines decision makers' selection of contextual control modes as described by Hollnagel's Contextual Control Model, and evaluates real-time, unobtrusive measures of a decision maker's immediate mode. In a two-part experiment, participants performed airline rescheduling tasks. The first portion varied task time limits, the second introduced a sudden change in the task. Participants reported operating in, and transitioning between, different contextual control modes in response to time limits and task changes. Computer interaction did not correlate to contextual control modes. Contextual control modes did not correlate with TLX ratings of demand and effort, but did correlate with TLX-frustration and TLX-performance ratings. The results suggest that decision making performance may be determined by use of context-appropriate contextual control modes, and imply that the design of decision aids should work to support those modes.

Introduction

Airline managers of a typical large U.S. airline are responsible for the daily operation of large regions or fleets of aircraft, often with 40-50 flights departing every hour. They oversee daily operations that are often disrupted by weather, ATC delays and maintenance problems. and are responsible for implementing flight delays, cancellations, "aircraft swaps" and the use of reserve crews to minimize the impact of such disruptions relative to the nominal flight schedule. Decisions must often be made quickly, frequently based on uncertain information. Many elements must be requested from other personnel (e.g., the maintenance department's estimate of a repair time). Other information must be retrieved from cumbersome text-based interfaces presenting data about hundreds of flights.

Our own observations have revealed that managers' approaches to this task can vary wildly. On a day with few disruptions the manager may consider many possible alternatives to minimize flight delays. Alternatively, on a busy travel day with major disruptions, the manager may resort to sweeping measures such as operating the entire fleet an hour behind schedule. This study hypothesized that these changes in decision making behavior may be described by different contextual control modes (CCM).

A large number of decision models which view decision making as the cognitive task of selecting from a set of alternatives. One accounting for some of the multiple decision models has recognized the tendency for human decision makers to "select" or "switch" cognitive strategies as a coping strategy in the face of stressors. Strategy switches include speed/accuracy trade-offs, task shedding, and the use of simpler strategies (e.g. Svenson, et al., 1993; Maule, 1997; Orasnau, 1997), which are not always explained simply as methods to reduce workload. While the selection of a strategy is often modeled as a cost-benefit activity (Maule, 1997), studies have also described cases where decision makers chose to increase their effort to maintain performance under perceived time constraints (e.g. Todd, et al., 1994; Kerstholt, 1996).

Hollnagel contends that the "the degree of control a person will have over a situation can vary. It seems reasonable to think of control as a continuous dimension where at one end there will be a high degree of control and at the other there will be little or no control" (Hollnagel, 1993). To better describe this continuum of control, Hollnagel has developed a classification of four contextual control modes:

• "*Scrambled control* denotes the case where the choice of next action is completely unpredictable or random." (Hollnagel, 1993, pp. 168)

• "*Opportunistic control* corresponds to the case when the next action is chosen from the current context alone, and mainly based on the salient features." (Hollnagel, 1993, pp. 169-170)

• "*Tactical control* is characteristic of situations where the person's event horizon goes beyond the dominant needs of the present, but the possible actions considered are still very much related to the immediate extrapolations from the context." (Hollnagel, 1993, pp. 170)

• "Strategic control means that the person is using a wider event horizon and looking ahead at higher level goals... The strategic control mode should provide a more efficient and robust performance, and thus be the ideal to strive for."(Hollnagel, 1993, pp. 170)

An important aspect of Hollnagel's contextual control model (COCOM) is the idea that individuals will transition between CCM to maintain control over a changing situation (Stanton, et al., 2001; Jobidon, et al., 2004). Hollnagel states that, "The change between control modes is determined by a combination of situational and person (or internal) conditions – in other words by the existing context…," (Hollnagel, 1993 pp. 194). Thus, the control mode must be appropriate to the context. An erroneous assessment of context, such as an incorrect subjective assessment of available time, may lead to use of a CCM that will not result in the best performance possible in the available time. For example, the impact of time pressure has been experimentally linked to CCMs in dynamic tasks, (e.g., Jobidon, et al., 2004) who concluded that time pressure and corresponding 'worse' CCMs lead to poorer performance.

However, the degradation in performance may not directly relate to choosing a 'worse' CCM. Inappropriate use of a higher control mode may also result in lower performance. For example, empirical studies by Oransanu et al. (1993) and Johnson et al. (2002) described how mismatches between context and decision strategies could have detrimental effects on performance. Unexpectedly, these mismatches can occur with reductions in workload, suggesting that CCMs and their appropriateness to the context can be better predictors of decision making performance than workload measures alone.

Decision support tools may be tailored for specific decision modes (Niwa, et al., 2002; Johnson, et al., 2002). However, very little work has been done to investigate measures which would allow real-time identification of an individuals'. Therefore, this research also investigates potential easily observable indicators of a decision maker's immediate CCM.

Objectives

The objectives of this study were twofold. First, we endeavored to verify the impact of time constraints and changes in task demands on human cognitive behavior as described by CCMs. Second, we sought to identify measures of CCMs including measures of information seeking behavior and a self-assessment.

Method

Participants

Participants in this experiment were undergraduate students. Data from 16 participants (12 males and 5 females) will be discussed here. The participants had a mean age of 22 years (ranging from 18 to 34 years), and had no previous airline scheduling experience. No selection criteria were used to qualify or disqualify participants.

Experiment Task and Procedure

Participants were asked to assume the role of airline manager for a small airline (4 airports, 4 aircraft and 12-16 flights). In the first part of the experiment the participants were presented with a disruption to an established flight schedule. Disruptions included weather and unexpected maintenance issues. They were instructed to strand as few passengers as possible while following some basic rules (e.g., all flights must terminate by midnight), and asked to find the best solution possible within a given time limit.

In the second part of the experiment, in addition to a the up-front disruption, a change in context was suddenly introduced part way into the task by telling participants that an aircraft had just announced they needed to divert to an airport due to a bomb threat, creating a further disruption. At the end of each run, participants were asked to record their solutions and the number of passengers it stranded. They were also asked to provide a self-assessment of workload and CCM.

The participants had access to complementary computer based and non-computer based information about the flight schedules. The information external to the computer mimicked information which is normally requested from a person who is not in the immediate vicinity, and thus carried a time cost. This external information represents information beneficial but not necessary for the completion of the task; by assigning a time cost to this supplementary information, its access suggests adequate subjectively available time for a tactical or strategic CCM.

Each participant conducted six runs. The first, a training run, had a simplified task to introduce the task, computer interface, and information available. The following five runs asked participants to find the best solution possible for a specified disruption in the time provided.

Apparatus

The experiment was conducted at a standard computer terminal with keyboard and mouse. The display was 17in. flat panel display set to a resolution of 1280 by 1024 pixels. Participants were also given a piece of paper and a pencil. The interface approximated the text-based terminal windows used by airline sector managers, with command buttons substituted for text-based commands.

Experiment Design

The two independent variables were time limit and the introduction of contextual change. In the first part of the experiment four time limits were used: 18, 13, 8, and 3 minutes. The final run (i.e., the second part of the experiment) introduced contextual change two minutes into the.

The scenario order, time limits and run order were balanced using a Latin square to minimize order, learning, and scenario effects. In the second part, the time limit was fixed at eight minutes, contained the same scenario task, and was always given last so that participants would not anticipate such a disruption in subsequent runs.

Dependent Measures

The data of interest were categorized into the following six groups:

Computer Interaction Key logging and mouse tracking software automatically recorded the frequency of requests for information from the computer and delete key hits.

Interaction External to the Computer External interaction was measured by the number of times the participant sought external information.

NASA Modified Task Load Index (TLX) Workload ratings were collected after each run via the six NASA TLX subjective rating sub-scales: mental demand, physical demand, temporal demand, performance, effort, and frustration.

Self-Assessment of Contextual Control Mode At the start of the experiment, subjects were briefed about the CCMs using Hollnagel's description for each. Then, on the questionnaire administered at the end of each run participants indicated the CCM they used during most of the task on a scale of 1-10, where the four CCMs were equally arranged and explicitly labeled at the 1 (scrambled), 4 (opportunistic), 7 (tactical) and 10 (strategic) marks. Additionally, participants were asked to state if they felt that they had transitioned from one CCM to another during the course of the task.

Performance Each scenario was designed to have at least four valid solutions. To standardize across all scenarios, the solutions were ranked according to the number of passengers stranded and the number of flights cancelled or delayed. The four best solutions were ranked one through four. All other valid solutions were given a rank of five. All invalid or incomplete solutions were assigned a rank of six.

Results

Experiment Part 1

A general linear model of the self-assessed CCM. This model indicated main effects due to scenarios ($F=3.989$, $p=0.024$) and time limit ($F=5.348$, $p=0.008$). Pairwise comparisons found differences between two scenarios ($p=0.017$). Time limit differences were found between 3min-13min ($p=0.017$), and 3min-18min ($p=0.007$) levels, as shown in Figure 1.

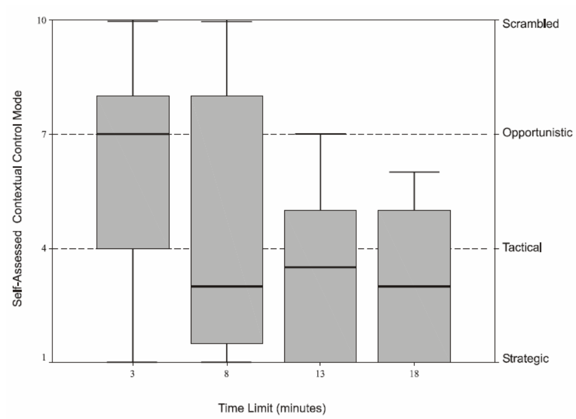


Figure 1. *Self-assessed CCM as a Factor of the Time Limit Imposed.*

A linear regression was performed on the reported CCM to examine the impact of observable indicators. The full model included the average time between mouse clicks, time limit, and the percentage of external information used. The model was found to be significant ($F=4.656$, $p=0.003$), however the average time between mouse clicks did not significantly contribute. There was a significant correlation between the percentage of external information used and time limit ($r=0.653$, $p < 0.001$). Likewise, a general linear model evaluated the six raw TLX subscale scores. Time limit was found to be a significant source of variance only in the TLX-temporal measures ($F=10.208$, $p<0.001$). Pairwise comparisons revealed that there were significant differences between the three minute level and all other levels ($p<0.05$).

A linear regression was performed on the raw TLX subscales. The full model included those measures which could be available for a real time assessment of CCM: the average time between mouse clicks, time limit, and the percentage of external information used. The model was found to be significant for the TLX-temporal subscale (F=9.736, p<0.001). Reduced models were found to be significant for the TLX-physical and TLX-frustration subscales. The reduced model for the TLX-temporal subscale only included time limit (F=28.976, p<0.001). The reduced model for the TLX-physical subscale included both time limit and the average time between mouse clicks (F=3.206, p=0.047), whereas the reduced model for the TLX-frustration subscale only included the average time between mouse clicks (F=6.111, p=0.016).

To compare self-assessed CCMs and workload, a linear regression was performed on the self-assessed CCMs, where the model included all six TLX subscales, average time between mouse clicks and the percentage of external information used (see Figure 2).

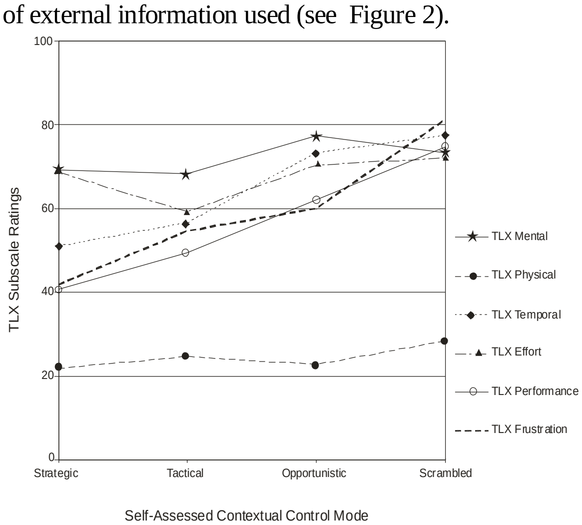


Figure 2. *TLX Subscale Scores by Self-Assessed Transition Direction in CCM*

The model was found to be significant (F=5.108, p<0.001). However, only the percentage of external information used, TLX-frustration, TLX-performance, and TLX-temporal subscales were found to significantly contribute to the model.

The effect of time limit, observable indicators, CCMs and TLX subscales on performance were then examined. A Kruskall-Wallis mean rank comparison found a marginally significant effect of time limit on participant performance (χ^2=6.333, p=0.096), as shown in Figure 3. Paired comparisons found a significant difference between performance in the 8 and 13 time limit levels (Z=-2.104, p=0.035).

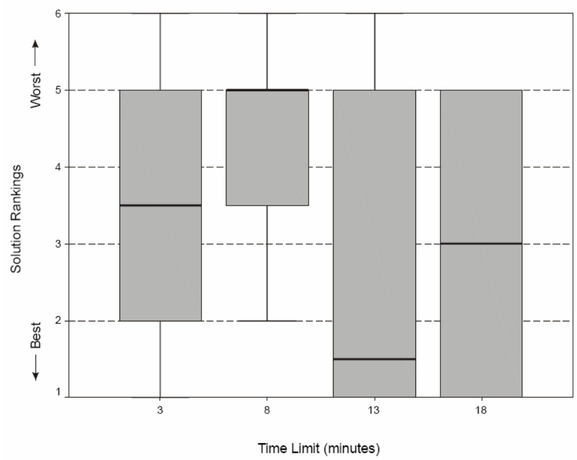


Figure 3: *Performance as a Factor of Time Limit*

A further Kruskall-Wallis mean rank comparison did not find differences in participant performance based on their self-assessed CCM. However, when individual paired comparisons were conducted a significant difference between participant performance was found between the opportunistic and the scrambled levels (p=0.033). A linear regression was performed on participant performance where the full model included all six TLX subscales. Neither the full model nor any of the individual TLX subscales were found to be statistically significant.

Experiment Part 2

In 10 of 16 participants (63% of the runs) there was a self-reported transition due to the contextual change of unexpectedly announcing (to the participant) that an aircraft was diverting to another airport, further disrupting the flight schedule. A general linear model was used to evaluate whether the inclusion of a contextual change affected the average time between mouse clicks, the TLX subscales, self-

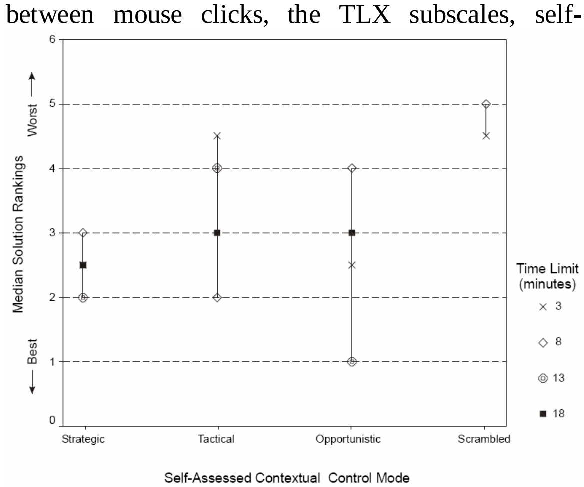


Figure 4. *Median Solution Performance by Self-Assessed CCM*

assessed CCM, or direction of CCM transition. Analysis of the model indicated that contextual change did not affect the average time between mouse clicks, the CCM or the CCM transition amount in a statistically significant manner. The analysis also indicated that the contextual change did affect the TLX-mental (F=11.309, p = 0.001), TLX-temporal (F=13.153, p=0.001) and TLX-frustration (F=4.681, p=0.034) subscales. Kruskall-Wallis mean rank comparisons found no significant effects due to the contextual change in performance, percentage of external information used or rule violations.

Adding the impact of a contextual change to the model generated in the first part of the experiment, which included time limit, the percentage of external information used, and contextual change, found that contextual change is also a statistical predictor of CCMs. The new model was significant (F=5.900, p<0.001.

Kruskall-Wallis mean rank comparisons were performed to see if the time limit, TLX subscales, or contextual change affected self-assessed CCM transitions. Of these, the only significant predictor of self-assessed CCMs is the TLX-frustration subscale. As shown in Figure 5, the TLX-frustration subscale was significantly affected by reported CCM transitions (^{2}=6.948, p=0.008), with a higher frustration level when participants reported a transition in either direction.

Discussion

The first part of the experiment examined the impact of time limits on human cognitive behavior as described by CCMs. The analysis revealed that, while there is a general trend for the self-assessed CCM to increase (become more strategic) with decreased time pressure, a linear trend is not strictly observed. Similarly, participants' performance did not linearly correlate with the self-assessed CCM. Many of the poorer performing data points correspond to self-assessments of 'opportunistic' control modes in the eight minute time limit condition and to 'tactical' control in the three and thirteen minute conditions, in addition to the conditions where the participants self-assessed their control mode as 'scrambled'.

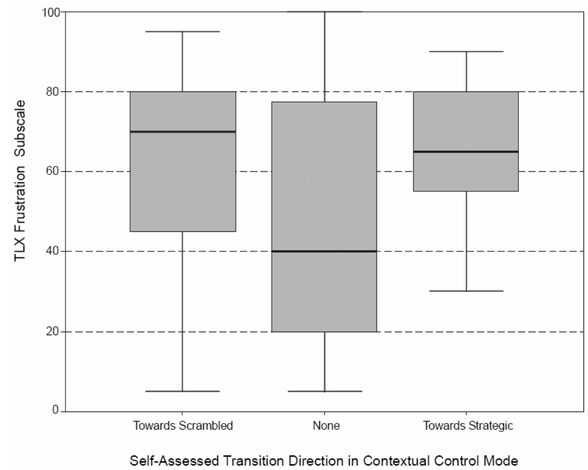


Figure 5. *TLX Frustration Scores by Self-Assessed Transition Direction in CCM*

These two findings may together correspond to the findings of the study by Johnson et al. (2002) in which participants sometimes appeared to ambitiously switch to inappropriate modes of behavior which could not generate high levels of performance within the time provided. These effects may correspond to poor assessments of subjectively available time in relation to the demands of the task.

The results also indicated that participants felt that their behaviors were more closely related to performance (and to frustration, defined in the TLX description as difficulties in achieving desired performance) than to measures of load and effort. As seen in Figure 2, only TLX workload ratings for performance and frustration were found to be statistical predictors of ratings of CCMs; TLX measures of demand and effort were not significant.

CCMs and workload differed in significant ways. Self-assessments of CCMs correlated with actual performance, whereas TLX ratings of perceived workload, including self-assessed performance, did not. Likewise, CCMs and the TLX subscales were predicted by different factors. For three of the TLX subscales, the observable indicators (average time between mouse clicks, amount of external information used, time limit) tested here were statistical predictors of TLX temporal by time limit, TLX physical by time limit and average time between mouse clicks, and TLX frustration by average time between mouse clicks. In contrast, CCM, while statistically predicted by the TLX performance and TLX frustration subscales, was not statistically predicted by any of the observable indicators.

In addition to the 'overall' CCM within each run, participants also reported transitioning between modes, with the transitions not statistically predicted by any of the observable indicators. Likewise, TLX-frustration was the only statistical predictor of transitions between CCMs during a run, albeit a comparatively weak predictor, as seen in Figure 5.

Conclusion

Participants in this study were able to provide a self-assessment of CCM. These self-assessments yielded a significant relationship to decision making performance and to contextual factors generally thought to impact performance, such as information sought, and self-assessed temporal demand, performance and frustration. These results support Hollnagel's representation of CCMs as involving more than a direct consequence of workload.

From the perspective of CCMs, the best performance within a given context (including time limit) will be attained when the decision maker applies the most appropriate CCM. Conversely, poor performance in this experiment corresponded not only to severe time limits demanding a 'lower' CCM, but also to perhaps over-optimistic attempts at 'higher' CCMs when sufficient time did not exist to carry them through. This perspective explains the results of earlier studies in which more time available sometimes led to a decrease in decision making performance.

These insights imply several design considerations. Decision makers operating within a fairly stable context might benefit from decision aids streamlined to support information-seeking, decision and action-taking behaviors which support the CCM most appropriate to that context. Keeping the context stable maybe seen as an important aspect of workload management. Evidence of this can be seen in standard ATC operating procedures. Controllers maintain focus on the near term and could be hypothesized as using tactical CCM, whereas the traffic flow managers are responsible for more strategic decisions and can by hypothesized to use tactical and strategic CCMs. When a controller is no longer able to manager the volume of traffic they are paired up with a D-side controller. This can be viewed as a controller no longer being able to operate at a tactical CCM, i.e. with out the additional controller they would be forced to operate at an opportunistic CCM due to traffic.

However, in many other aviation situations the decision maker's context can vary from hour to hour and from day to day, such as the airline rescheduling task examined here and other aviation related jobs. In these cases, the decision aid may need to be capable of supporting several different CCMs. This may be achieved through one large interface which centrally emphasizes the most salient information needed in opportunistic CCMs while also supporting the information seeking and explorative behaviors corresponding to tactical and strategic CCMs.

One could argue that the differences in assessment of how much information to give a pilot or a controller stems from CCM. Depending on which "level" the pilot or controller is operating at will greatly influence how much information and which types of displays would be most helpful. At an extreme, an aid may be envisioned with separate interfaces for each of the CCMs potentially employed by its user. Such an aid could, in theory, switch automatically between interfaces in response to its user's transitions between CCMs, i.e., an "adaptive decision aid" equivalent to "adaptive automation." However, as the real-time indicators examined in this study were not able to statistically predict CCM, some other indicators or methods of assessing the user's control behavior would be required. Participants' ability to self-assess their CCM suggests that decision makers may be able to manually switch between interfaces to obtain the level of support they require, i.e., an "adaptable decision aid" may be a better approach to support pilots and controllers by allowing them to chose how much information they need. With experience, interface switching may itself be another component of an expert's adaptation to the operating environment. Before such expertise is developed, however, another potential role of the interface may also be to present contextual factors that allow the controller, pilots, and airline operations managers to better select the CCM most appropriate to their immediate situation.

Acknowledgements

This material is based upon work supported under a National Science Foundation Graduate Research Fellowship, and the experiment was supported in part through grants awarded to Julie A. Jacko by the Intel Corporation and the National Science Foundation (BES-9896304)

References

Hollnagel, E. (1993). Human reliability analysis: Context and control. London, UK: Academic Press.

Hollnagel, E. (2002). Time and Time Again. Vol. 3(2), 143-158.

Hutchins, E. (1995). Cognition in the Wild.

Jobidon, M.-E., Rousseau, R., & Breton, R. (2004). Time in the Control of a Dynamic Environment. Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting, 557-561. HFES.

Johnson, K. E., Kuchar, J. K., & Oman, C. M.Experimental Study of Automation to Support Time-Critical Replanning Decisions. Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, 5. 2002: HFES.

Kerstholt, J. H. (1996). The Effect of Information Costs on Strategy Selection in Dynamic Tasks. Vol. 94, 273-290.

Maule, J. A. (1997). Strategies for Adapting to Time Pressure. R. Flin, E. Salas, M. Strub, & L. Martin, Decision Making Under Stress, 271-280. Aldershot, UK: Ashgate Publishing Ltd.

Niwa, Y., & Hollnagel, E. (2002). Principles of Performance Monitoring in Coupled Human-Machine Systems. Johannsen, IFAC Analysis, Design and Evaluation of Human-Machine Systems, 303-307.

Orasnau, J. (1997). Stress and Naturalistic Decision Making: Strengthening the Weak Links. R. Flin, E. Salas, M. Strub, & L. Martin, Decision Making Under Stress, 43-66. Aldershot, UK: Ashgate Publishing Ltd.

Svenson, O., & Edland, A. (1993). On Judgment and Decision Making Under Time Pressure and the Control Process Industries. IEEE International Conference on Systems, Man & Cybernetics, 367-375. IEEE.

Todd, P. A., & Benbasat, I. (1994). The Influence of Decision Aids on Choice Strategies Under Conditions of High Cognitive Load. Vol. 24(4), 537-547.

CHANGES IN SAFETY ATTITUDES IN A CANADIAN REGIONAL AIRLINE FOLLOWING A MERGER

Dr. Mark Fleming
Saint Mary's University
Halifax, Nova Scotia Canada

Bernadette Gatien
Saint Mary's University
Halifax, Nova Scotia Canada

Captain Rob Vacher
Air Canada Jazz
Halifax Nova Scotia Canada

The present study examines the impact a merger had on pilot safety attitudes. Pre and post merger safety attitudes among a sample of Canadian pilots were examined using the Flight Management Attitudes Questionnaire 2.1 (FMAQ) (Merritt, Helmreich, Wilhelm, & Sherman, 1996) and the Flight Management Attitudes and Safety Survey (FMASS) (Sexton, Wilhelm, Helmreich, Merritt, & Klinect, 2001). Data were collected from 232 airline pilots prior to a large-scale merger using the FMAQ 2.1. Approximately 1 year following the merger, FMASS data were collected from the newly merged organization. We hypothesized that pilots' safety attitudes were negatively impacted due to uncertainty and organisational change following the merger. Results of the study indicate that post merger attitudes were significantly different on the teamwork, job attitudes and safety culture facets of the scale, however there were no significant differences in the stress recognition following the merger. In addition, the psychometric properties of the FMASS were examined. Implications for this study include understanding how change within an aviation organisation impacts safety attitudes thus impacting the overall safety culture of the newly created organisation.

Introduction

Literature on organizational mergers indicates that the nature of a merger can have either positive or negative impact on employee attitudes. Specifically hostile mergers or takeovers often foreshadow a decrease in employee attitudes towards job characteristics, work relationship satisfaction job security, organizational commitment (Newman & Krystofiak, 1993).

Fairfield, Ogilvie and DelVecchio (2002) also found that employee attitudes towards the organization and their jobs become significantly more negative following a hostile merger than those employees involved in a friendly merger. Specifically employees involved in the hostile merger had lower attitudes towards organizational commitment and job satisfaction (Fairfield and Ogilvie, 2002). Limited research is available on the impact of organization mergers of employees' safety attitudes. The potential impact of shifting attitudes towards safety can have serious implications on both the organisation and individual employees. Negative safety cultures and attitudes have been linked to various fatal organizational disasters (e.g. Chernobyl, Piper Alpha). Safety culture or attitudes is a key concept in understanding the role organizations play in major accidents and disasters. It is essential that organizations recognize and appreciate the importance of having and maintaining a positive safety culture especially after undergoing a major organizational change.

The current research will attempt to address this current gap in the literature by examining the mean differences in safety attitudes before and after a corporate merger. The current merger situation was not considered a hostile one, however the merger did involve a great deal of organizational restructuring (physically) and uncertainty regarding, relocation of staff, and layoffs and company future. It was hypothesized that safety attitudes among all levels of pilots were negatively impacted given the level of uncertainty surrounding the impact of the merger.

Evaluation of CRM Attitudes

Attitudes are not overtly observable, and must, therefore, be measured using either indirect methods such as behavioural observation or direct methods such as surveys or questionnaires (Azjen, 1991). The Flight Management Attitudes Questionnaire (FMAQ) (Helmreich, Merritt, Sherman, Gregorich, & Wiener, 1993) is the most commonly used measure of CRM attitudes within the aviation industry. It is predominately used to evaluate the current status of safety attitudes and as a training evaluation tool (e.g., Salas, Fowlkes, Stout, & Milanovich, 1999). The original FMAQ was developed to specifically measure cockpit management attitudes and was therefore, referred to as the Cockpit Management Attitudes Questionnaire (CMAQ) (Helmreich, 1984). In 1988, a revised version of the CMAQ was developed because the existing version did not account for cross-cultural attitudes (Gregorich, Helmreich, & Wilhelm, 1990). The FMAQ was developed as an extension to the CMAQ, containing all of the original CMAQ items in addition to new items that were based on Hofstede's (1982) four dimensions of national culture (power distance, individualism, collectivism, uncertainty avoidance and masculinity-femininity) (Helmreich & Merritt, 1998). The original version of the FMAQ contained 82 Likert scale items, designed to measure pilot attitudes towards command, communication, stress, rules, automation, organisational climate and work values (Helmreich & Merritt, 1998). The questionnaire has since been revised including the FMAQ 2.0 international version and the FMAQ 2.1 USA/Anglo version and the Flight Management Attitudes and Safety Survey (FMASS). The FMASS was developed out of a need for a short version of the FMAQ (Sexton et al., 2001).

The FMASS contains four factors: safety culture, which is defined as "the extent to which individuals perceive a genuine and proactive commitment to safety by their organisation" (Sexton et al., 2001). The second factor is job attitudes, which are defined as "the level of satisfaction with the organisation and the individual's reactions to his or her job experience". Teamwork is the third factor and is defined as "the level of satisfaction with the quality of teamwork and cooperation experienced with other crew members, gate agents, ramp personnel, flight attendants, dispatch, maintenance, and crew scheduling" (Sexton et al., 2001). The final factor is termed stress recognition and is defined as "the extent to which individuals acknowledge personal vulnerability to stressors such as fatigue, personal problems and emergency situations" (Sexton et al., 2001, p. 5-9).

Methodology

Participants

Pre-merger. A total of 232 pilots voluntarily completed the FMAQ 2.1 (Helmreich et al., 1993) following Crew Resource Management (CRM) training sessions. The response rate for the pre-merger was 90%. This sample contained only two bases located in one region of Canada.

Post merger. Following the merger all pilots from the newly created organization were surveyed following the merger. A total of 726 pilots voluntarily completed the FMASS following CRM training sessions, for a response rate of 75%. In total 204 participants from the original two bases of the pre-merger organization were surveyed in the second survey. Due to the fact that the questionnaires were completed anonymously it was not possible to link individual responses in the pre-merger survey to their responses in the post merger survey. Fortunately there were limited changes in personnel at the pre-merger locations, so it was possible to test for differences at the base level.

Measures

Pre-merger data was collected using the FMAQ 2.1 USA/Anglo version (Helmreich et al., 1993). Post merger data was collected using the FMASS (Sexton et al., 2001). Two different versions of questionnaires were used in the data collection however, only those items which were identical in both versions were used in comparing pre and post data. All of the items for the four factors that make up the FMASS were included in the FMAQ, therefore it was possible to look at differences in factor scores. All other items were excluded from the analysis. See table 1 for a list of items

Table 1. *Factors and Items*

Safety Culture
The managers in flight ops listen to us, and care about our concerns
I am encouraged by my supervisors and co-workers to report unsafe conditions
Management will never compromise safety concerns for profitability
My suggestion about safety would be acted upon if I expressed them to management
Job Attitudes
Working here is like being part of a large family
I like my job
Pilot morale is high
I am proud to work for this organisation
Pilots trust senior management at this airline
Stress Recognition
Personal problems can adversely affect my performance
My decision making abilities are as good in emergency situations as routine flying conditions

Table 1. *Factors and Items*

	I am more likely to make judgment errors in an emergency
	My performance is not adversely affected when I am working with a less experienced or less capable crew member
	A truly professional crewmember can leave personal problems behind when flying
Teamwork Scale	
	Describe your personal perception of the quality of teamwork and cooperation you have experienced with the following
	Other cockpit crewmembers
	Flight Attendants
	Dispatch
	Crew Scheduling
	Maintenance

Data Analyses

Using Statistical Package for the Social Sciences (SPSS, 1999), the data were cleaned by examining minimum and maximum response values, ranges, means, standard deviations, skewness, kurtosis, and standardized scores. The data was also screened for univariate and multivariate outliers in addition to violations of assumptions.

Pre and post merger differences were examined using a Mann Whitney U non parametric test. This analysis was chosen due to sample limitations.

Results

Results of the Mann Whitney U tests indicate that there was significant difference in attitudes following the merger for three of the four FMASS factors. Specifically, safety culture attitudes became more negative ($Z = -8.00$, $p < .001$). Attitudes towards the quality and teamwork and cooperation also became more negative ($Z = -6.44$, $p < .001$). Similarly job attitudes also became significantly more negative following the merger, ($Z = -11.22$, $p < .001$). No significance difference in attitudes on the stress recognition scale (see figure 1). Reliabilities for each of the scales ranged from good .79 to poor .37 (see table 2 reliability coefficients).

Table 2. *Reliability coefficients*

Scale	**Alpha**
Pre-merger	
Teamwork	.77
Safety culture	.78
Job Attitudes	.79
Stress Recognition	.53
Post merger	
Teamwork	.69
Safety culture	.37
Job Attitudes	.77
Stress Recognition	.57

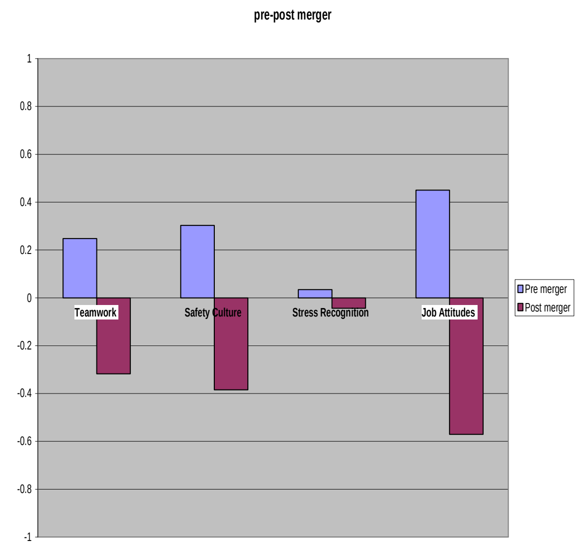


Figure 1. *Difference between pre and post merger.*

Discussion

The Airline industry currently faces significant challenges, with a number of major carriers either just emerging from bankruptcy protection or still in bankruptcy protection. To remain solvent companies have made pilots redundant and reduced salaries and benefits for those who remain. In some instances there has been uncertainty about pension entitlement. These challenges are likely to have an impact on pilot attitudes. Although the importance pilots place on safety is unlikely to change, their perceptions about management priorities is likely to be adversely effected.

The current study investigated the impact of a company merger on pilot attitudes. There was a significant mean difference in pilot attitudes following the merger. Specifically they were less positive about the safety culture, job attitudes and teamwork. This difference in pilot attitudes suggests that organizational change and uncertainty can negatively impact pilot perceptions. It is therefore important to consider the mechanisms by which a merger can have such an impact.

The safety culture dimension measures pilots' perceptions about commitment to safety within the organization. It is likely that during a merger or any organizational change people focus on issues relating to the change, such as organizational structure and staffing. How people spend their time and what they speak about are two important indicators that people use to judge commitment to safety. For example if a manager never speaks about safety and dedicates their time solely to profit maximization then people will assume that safety is not a priority for them. During a merger it is likely that managers' time and energy will be focused on issues related to the change. The additional demand placed on managers' time means that are likely to have less time to devote to other issues, including safety. This does not mean that the managers think safety is unimportant, but when faced with other immediate demands their commitment to safety may become less visible. This is especially likely to happen in aviation since safety related incidents are relatively infrequent. It is therefore important for managers to be cognizant of this risk, so that they visibly demonstrate their commitment to safety to maintain a positive safety culture during a period of change.

Similarly, it is likely that pilots' satisfaction with the organization was adversely effected by the merger due to uncertainty and changes in management behavior. The loss of their previous corporate identity may also have had a negative impact. In general, mergers involve changes in the management structure, personnel and approach to staff management. In addition, mergers are often undertaken to achieve cost savings through personnel reduction. Therefore mergers are often associated with concerns about job security. It is therefore not surprising that pilot attitudes towards their organization were negatively affected by the merger. Although it is difficult for organizations to prevent pilots from being concerned about their job security, they should attempt to mitigate the negative impact by trying to maintain positive management-pilot relations during the period of uncertainty. This involves managers (who may be concerned for their own job) proactively working to maintain positive relations. Interventions could include holding open meetings where pilots can openly discuss concerns, and managers meeting with pilots in neutral settings (social events) to maintain an open dialogue. It may also be beneficial to introduce joint pilot-manager training programs, to promote open dialogue and an understanding of the challenges being faced by each group.

The negative impact that the merger had on team work is likely to be due to changes in personnel as the merger will have introduced team members from the previous organization. In addition, it is likely that the negative impact on culture and job attitudes will have had a knock on effect on perceptions of team work. For example, if dispatchers are also less satisfied with the organization, then this may have a negative impact on their performance.

Limitations & Future Research

The limitations with the current sample are the low level of reliability for the stress recognition scale prior to and following the merger. Additionally the reliability for the safety culture scale dropped significantly following the merger. Another limitation of this study is the inability to match individual pilot responses from the pre-merger sample with the post merger sample. Therefore the pre and post merger groups were treated as independent samples, yet in reality they were the same individuals, this violates the assumptions of independence required by many statistical procedures. This limitation was mitigated by using a very conservative statistical test. Future research should examine the psychometric properties of each of the individual subscales of the FMASS as well as the reliability for the entire scale. Additionally future research should examine whether there is a lasting impact of mergers on pilot safety attitudes.

References

Azjen, I. (1991). The theory of planned behaviour. *Organizational Behavior and Human Decision Processes, 50,* 179-211.

Fairfield-Sonn, J.W., Ogilvie, J.R., & DelVicchio, G. A. (2002). Mergers, acquisitions and long term employees attitudes. *Journal of Business & Economic Studies, 8*, 1-16.

Helmreich, R. L. (1984). Cockpit management attitudes. *Human Factors, 26,* 583-589.

Helmreich, R. L., Merritt, A. C., Sherman, P. J., Gregorich, S. E., & Wiener, E. L. (1993). The Flight Management Attitudes Questionnaire (FMAQ). NASA/UT/FAA Technical Report 93-4. Austin, TX: The University of Texas. [FMAQ 1.0 introduction 1993.pdf].

Newman, J.M., Krzystofiak, F. J. (1993). Changes in employee attitudes after an acquisition. *Group & Organization Management, 18,* 390-410.

Salas, E., Fowlkes, J. E., Stout, R. J., Milanovich, D. M., & Prince, C. (1999). Does CRM training improve teamwork skills in the cockpit? Two evaluation studies. *Human Factors, 41,* 326-343.

Sexton, J. B., Wilhelm, J.A., Helmreich, R.L., Merritt, A.C., & Klinect, J.R. (2001). Flight management attitudes and safety survey (FMASS). A short version of the FMAQ. University of Texas Human Factors Research Project Technical Report 01-01. [Safety Culture OSU 2001.pdf].

EXPERIMENTAL ANALYSIS OF THE INTEGRATION OF MIXED SURVEILLANCE FREQUENCY INTO OCEANIC ATC OPERATIONS

Laura Major Forest[1] & **R. John Hansman**[2]
[1]C.S. Draper Laboratory, Cambridge, MA 02139 USA; lforest@draper.com
[2]International Center for Air Transportation, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA; rjhans@mit.edu

Technical capabilities for improved surveillance over the oceans are currently available through the use of satellites. However, all aircraft operators will not equip simultaneously because of the high costs required. Consequently, as these CNS systems are integrated into oceanic air transportation architecture, the controller will have to manage the current low frequency surveillance in parallel with enhanced surveillance. The cognitive effects of the mixed equipage environment were studied through experimental analysis. The results confirm that there are human performance issues with integrating mixed surveillance capabilities, which may result in safety and efficiency limitations.

Introduction

The effects on human performance capabilities of integrating position information from dissimilar sources, with significant differences in update rate and reliability, have not previously been addressed. However, there are plans for such integration into oceanic air traffic control (ATC) operations in the near future (Federal Aviation Administration, 2002).

Oceanic ATC Surveillance

Since air traffic over the oceans is out of radar coverage surveillance presently consists of pilots reporting their position over high frequency (HF) radio at designated waypoints, which occur approximately once every hour (Civil Aviation Administration, 2002). All HF communication is conducted through a third-party communication relay service (e.g., ARINC in the United States). This indirect surveillance process is shown in Figure 1. HF radio is unreliable due to the interference of solar storms and other anomalies. The above limitations result in a high amount of latency and unreliability associated with the current surveillance process.

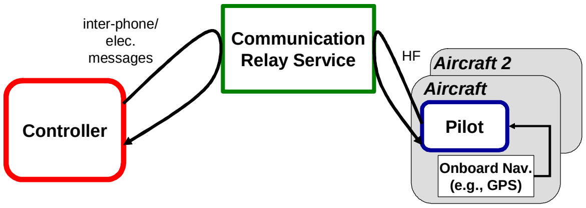


Figure 1: *Current surveillance is conducted by pilots reporting their position, over HF radio, through a third party communication relay service.*

Traditionally the position reports have been displayed to the controller through the use of paper flight strips. The strips are organized into columns, with each column representing a position reporting point. This allows the controllers to monitor the aircraft by comparing time at waypoint. Currently the majority of oceanic traffic follows standard routings that are usually deconflicted. Therefore, in the present system the controllers nominally ensure separation at the waypoints and assume separation minima will be maintained in between.

Satellites have introduced the opportunity for improved surveillance. One such opportunity is provided by addressable automatic dependent surveillance (ADS-A)[1]. ADS-A automatically sends flight information, through a satellite communication link, to specified addressees (typically ATC ground stations) at specified intervals. The intervals are determined by contracts between the aircraft operators and the ATC centers. The ADS-A surveillance process, shown in Figure 2, significantly improves the frequency of surveillance updates and increases reliability.

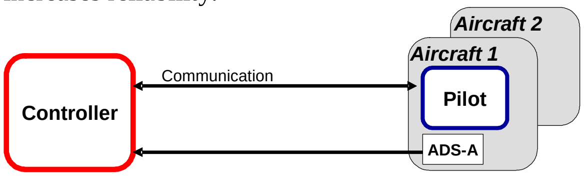


Figure 2: *Future oceanic surveillance involves ADS-A reports through a satellite communication link.*

For ADS-A capability aircraft operators must retrofit their aircraft with onboard avionics equipage. The high cost of the equipage restricts many aircraft

[1] ADS-A is referred to as ADS-Contract (C) in Europe.

operators from acquiring this new technology. Consequently, oceanic airspace will consist of mixed surveillance capabilities in the near-term.

Spatial situation displays are being integrated into oceanic ATC workstations to take advantage of ADS-A and other opportunities for improved surveillance and decision support. Currently, aircraft surveilled by pilot position reports are represented on the spatial display. Their position is represented as a continuous projected path based on an extrapolation of the filed flight plan, any changes made to the flight plan, and computer models. Once ADS-A surveillance information is integrated with the pilot position reports, the controllers may have difficulty distinguishing between the aircraft equipped with high and low frequency surveillance. This may cause the controllers to treat all aircraft as if they were equipped with low frequency surveillance (lowest common denominator effect).

Based on field studies of current operations and human-centered systems analysis, other human factors issues are hypothesized to emerge when mixed surveillance capabilities are integrated into common airspace. These issues are increased controller workload, decrease in situation awareness, and the possibility that controllers will choose to maneuver aircraft equipped with high frequency surveillance when in conflict with unequipped aircraft, which will negate the advantages of equipping. The present study investigated these issues further through a part-task experimental analysis.

Experimental Analysis

For this study air traffic controller trainees ran three scenarios on a PC-based simulator. The scenarios were modeled after the current, near-term, and far-term future oceanic operations.

Participants

The participants for the experiment were nine air traffic controller trainees. The experiment took place approximately three weeks prior to their full ATC certification. A questionnaire was administered to determine the level of participants' operational control experience. As part of their air traffic controller training the participants controlled in the ATC operational environment under the supervision of fully certified controllers for an average of 24 months (SD=0.899). Their experience was in Enroute Centers and Approach Centers (TRACON and Tower).

Air Traffic Control Simulator

A PC-based low fidelity ATC simulator was developed at MIT for this experiment. As shown in Figure 3, generic oceanic airspace was simulated and displayed through a spatial representation. This display was modeled after the spatial displays currently used at oceanic ATC facilities.

The display consists of aircraft targets, datablocks, jet routes, and fixes. A circle with a radius equal to the minimum separation surrounds each aircraft target. The circle can be removed during the simulation by right clicking the aircraft icon. The datablock includes the aircraft callsign, equipage information (ADS or non-ADS), altitude, and speed. The aircraft position on the spatial display is updated once per surveillance update. Consequently the variance in surveillance update rates, during the mixed equipage scenario, is reflected on the traffic display.

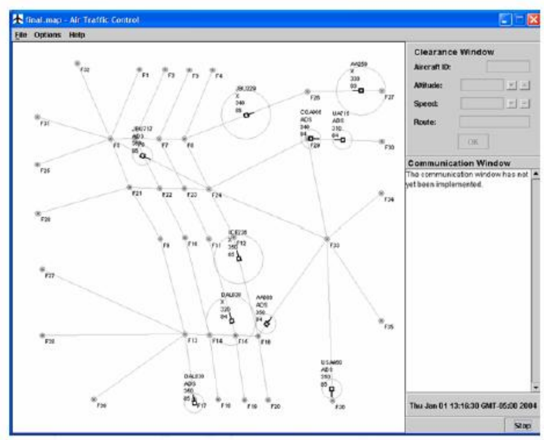


Figure 3: *PC-based simulator used for the experimental analysis*.

Experiment Design

The participants were presented with three, five to seven minute scenarios. For each of these scenarios there was moderate traffic and the airspace geometry was varied, however the level of complexity was held constant. Four conflicts were built into each of the three scenarios in random order. The conflicts were changed in a superficial manner to maintain consistency across the scenarios, but not result in a "training effect". These conflicts included two merging conflicts between two aircraft, a head on conflict, and a more complex conflict, involving four aircraft that were all converging at one point. There were also three pilot requests, which were easy to medium difficulty. The responses to the pilot requests were not used in the analysis.

Independent Variables

The independent variables were surveillance frequency (low, high, and mixed) and separation minima (modeled after current and future operations). The three scenarios are described in Table 1. One of the scenarios consisted of aircraft equipped with low frequency surveillance (1 update per 30s) and separation minima of 50 nm. In another scenario all of the aircraft were equipped with high frequency surveillance (1 update per second) and separation minima was reduced to 20 nm. The third scenario consisted of a 50% mix in surveillance equipage. In this scenario separation minima was reduced to 20 nm only for aircraft equipped with high frequency surveillance.

Table 1: *Design of the three scenarios.*

Surveillance Frequency	Separation Minima
High (1 upt/s)	20 nm
Low (1 upt/ 30s)	50 nm
Mixed	Mixed

Dependent Variables The dependent variables focused on the human performance impact on the controller since benefits to the system rely on the capability of the controller to implement the reduced separation safely. The first variable of interest was workload. Because of time constraints, approaches to measuring workload that would lengthen the experiment time, such as NASA-TLX or a secondary task, were not possible. Instead a subjective rating on an anchored five point scale by the participants of the difficulty of each scenario was used in addition to a difficulty ranking of the three scenarios. Another dependent variable was situation awareness, which was measured by use of the performance-based testable response method (Pritchett, 1996). Controller trust is vital for acceptance and implementation of changes to the ATC system. Trust was evaluated by the participants rating their confidence in aircraft position information for aircraft with high and low surveillance frequencies on an anchored five point scale. The final dependent variable was the surveillance type of the aircraft the subject maneuvered during a mixed conflict. This was evaluated only during the mixed scenario. It was hypothesized that the participants would choose to maneuver aircraft equipped with high frequency surveillance when faced with a mixed conflict.

Results

Each of the dependent variables was analyzed using a one way ANOVA. The ANOVA analysis was used to test for statistical significance in the difference in the means of the three scenarios. To further compare the scenarios with mixed and high frequency surveillance a t-test was used. The Bonferoni correction was used to adjust for the additional test.

Scenario Difficulty

The results from the difficulty rating showed a significant positive effect of surveillance frequency, F(2, 8)=4.795, p=.018. Using the related-pairs t-test, a significant difference was identified between the high frequency and mixed scenario pair, p=.002. These results, shown in Figure 4, demonstrate that integrating high and low frequency surveillance does not result in an improvement in workload from the current operations.

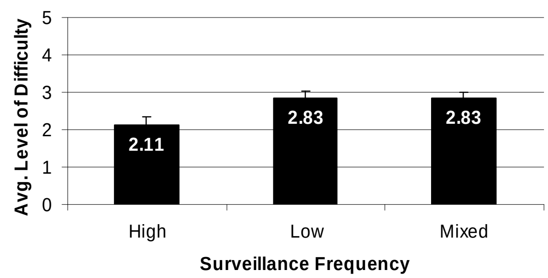


Figure 4: *Results from participant rating of the difficulty of each scenario on an anchored scale of 1 to 5, with 5 being the most difficult. Standard error bars are given.*

The post-experiment ranking of the difficulty of the three scenarios also revealed a significant effect consistent with the post-scenario ranking, F(2, 8)=7.44, p=.003. As shown in Figure 5, 67% of the participants found the mixed scenario to be the most difficult.

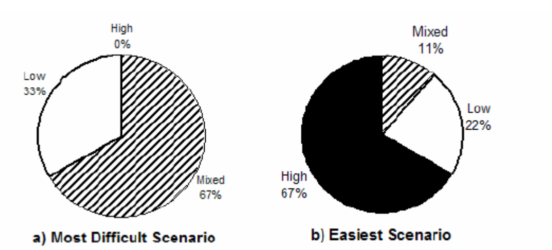


Figure 5: *Results from participant ranking of the three scenarios.*

Situation Awareness

Situation was measured by scripting four conflicts and measuring performance, to determine if participants were aware of the conflicts and how quickly they recognized and resolved them. There was a non-significant increasing trend in the number of conflicts resolved. Low significance was expected because of the conservative culture amongst air traffic controllers. Priority on resolving conflicts is valued much more than efficiency.

There was also a non-significant trend in the time to recognize the four scripted conflicts in the three scenarios, $F(2, 8)=2.400$, $p=.115$. The trend can be seen in Figure 6. There was not a significant trend evident in the time required to resolve the conflicts.

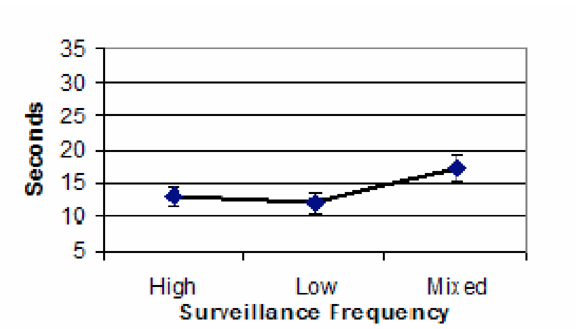


Figure 6: *The average time it took the participants to recognize the four scripted conflicts in the three scenarios. Standard error bars are given.*

Participant Confidence

There was a significant effect of surveillance frequency on confidence, $F(2,8)=21.951$, $p=.002$. As expected, the subjects rated their confidence in the position of aircraft with high frequency surveillance much higher than their confidence in that of aircraft with low frequency surveillance. The results are demonstrated in Figure 7.

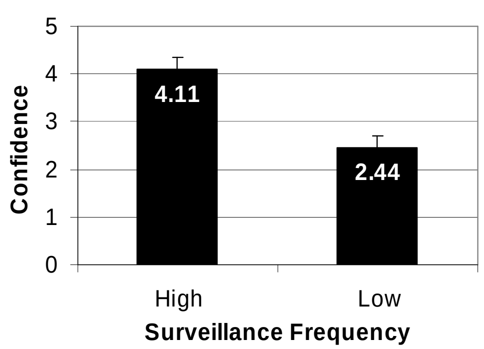


Figure 7: *Participant rating of their confidence in the position information for aircraft with high and low frequency surveillance on an anchored scale of 1 to 5, with 5 being "Very Confident".*

Aircraft Maneuvered

During the post-experiment survey, participants were asked which aircraft they were more likely to maneuver in a mixed conflict, aircraft equipped with high or low frequency surveillance. All nine participants responded they would be more likely to maneuver aircraft equipped with high frequency surveillance. This result matches the trend in their performance during the mixed scenario. A significant difference was found between the number of high frequency and low frequency aircraft chosen to maneuver by the participants, $F(1,8)=20.455$, $p=.0003$. The number of high frequency and low frequency aircraft that each participant chose to maneuver to resolve the four conflicts in the mixed scenario is shown in Figure 8. Some participants did not resolve all four conflicts because some of the conflicts were missed or averted with a previous maneuver.

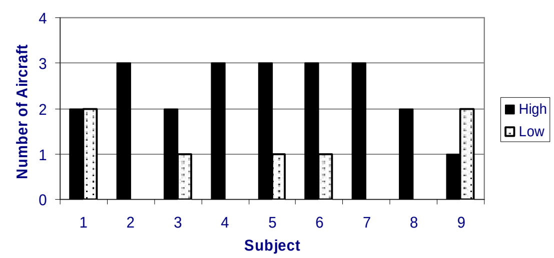


Figure 8: *The number of aircraft each subject maneuvered to resolve the four scripted conflicts in the mixed scenario.*

This result was expected based on the controller rating of their confidence in position information for aircraft equipped with high and low frequency surveillance. Maneuvering aircraft with frequently updated position information creates a more robust resolution to a mixed conflict.

Conclusions

The results from the part-task experiment confirm the hypothesis that controller cognitive limitations will negatively impact the advantages achieved by integrating aircraft equipped with improved surveillance into oceanic ATC operations. Safety may be compromised due to a potential increase in controller workload and degradation in situation awareness. The efficiency benefits associated with improved surveillance also may not be achieved because controllers will nominally maneuver aircraft equipped with the highest frequency surveillance, taking the aircraft off of their planned path, to resolve mixed conflicts.

Airspace segregation and display support are proposed to alleviate the human performance costs associated with the mixed surveillance environment. Airspace segregation reduces the complexity for the controller by removing the need to apply different strategies based on individual aircraft capabilities. Each airspace region will have a set of required equipage capabilities associated with the region. This allows the controller to apply the same procedures and control strategies within each airspace region. Airspace segregation is currently used for aircraft equipped with reduced vertical separation minima (RVSM) and required navigation (RNP). The majority of flight levels over the oceans are dedicated to RVSM equipped aircraft. Standard routings, such as the oceanic track structure (OTS), are dedicated to aircraft equipped with RNP-10[2].

Further research needs to focus on strategies for segregated operations and the display support required to support these operations. These strategies need to be consistent with the RNP concept, since the concept is included in plans for future reductions in separation minima. Additional studies into how to display various surveillance frequencies on a single display are also needed.

References

Civil Aviation Administration Iceland (2002). *Manual of Operation*. Reykjavik.

Federal Aviation Administration (2002). Strategic plan for oceanic airspace enhancements and separation reductions. Technical Report Draft 1.0. Washington D.C..

Gordon-Smith, M. (2003). RNP promises significant economic benefits to the airline industry. *ICAO Journal*, 58(3):23–25, 33–34.

Pritchett, A. R., Hansman, R. J., & Johnson, E. N. (1996). Use of testable responses for performance-based measurement of situation awareness. *Proceedings of International Conference on Experimental Analysis and Measurement of Situation Awareness*. Daytona Beach.

[2] Aircraft certified for RNP-10 must maintain a cross-track and along-track navigational accuracy of 10 nm, 95% of the time (Gordon-Smith, 2003).

AN EVALUATION OF HEAD-MOUNTED DISPLAYS IN AN AIRBORNE COMMAND AND CONTROL SIMULATION ENVIRONMENT

Scott M. Galster[1], **Rebecca D. Brown**[2], **Brent T. Miller**[1], **Alison M.Tollner**[3]

[1]Air Force Research Lab
WPAFB, OH, USA

[2]General Dynamics, AIS.
Dayton, OH USA

[3]University of Cincinnati
Cincinnati, OH, USA

An examination of HMDs to ameliorate the problems associated with display clutter in an Air Battle Management environment was conducted. Information to complete tasks was given via two HMDs, on the primary display, or via paper. The results indicated that the paper condition engendered a higher percent of correct responses and faster response times in several of the tasks performed. The specific experimental results are presented and future experimental design propositions are discussed.

Introduction

Air Battle Management (ABM) is a complex and demanding task that requires operators to direct the implementation of a dynamic air tasking order (ATO) and control the tactical execution of air-to-air and air-to-ground operations specified by that ATO. In order to do so, they typically monitor and manipulate a situation display (SD) comprising a map overlaid with landmarks, geographical features, and moving tracks representing the air and ground assets of coalition and enemy forces, as well as neutral tracks and those tracks for which positive identification is lacking. In addition to this surveillance component, ABMs perform a myriad of secondary tasks, such as associating coalition assets with targets, coordinating air-to-air refueling, and responding to alarms and alerts. Those secondary tasks often require a portion of the visual display that is occupied by the SD, and typically occlude part or all of the SD, which can potentially lead to decreased performance on one or more of the concurrent tasks.

One solution that addresses the problem of display occlusion and clutter is to increase the size of the display area by adding additional or larger monitors. Although this solution does not require the SD to be occluded, it does require overt shifts of visual attention, potentially involving head or eye movements. In this design scenario, information would be placed further away from the center of the workstation not only resulting in time spent looking away from the SD, it would also require additional time to reacquire the situation once the off-axis task has been dealt with. Furthermore, this solution is not practical in many ABM environments, especially those sited on airborne platforms, due to space limitations.

Most modern ABM workstations are transitioning to electronic documentation of information and abandoning traditional paper manuals therefore creating an ever increasing need for the display of battle space information. The need for increased screen space in command and control environments has been expressed (St. John, Manes, Oonk, & Ko, 1999). The increasing need for the display of information along with space limitations bring about questions of how information can be displayed in a manner that is space efficient, useful and least disruptive to other tasks.

Technological advancements in processing speed and the miniaturization of technology have led to several possible alternative solutions. One potential solution for improving the problem of display occlusion is to use a head-mounted display (HMD) to provide additional screen space. HMDs have received considerable attention and investigation due to their ability to enhance human perception and performance in certain complex work environments. HMDs have been used successfully in various environments including surgery, entertainment, manufacturing, military applications, training, and education. For instance, HMDs have been proven to enhance the operational effectiveness of Apache AH-64 helicopter pilots (Stelle, Reynolds, Rash, Peterson, & Leduc, 2003) as well as provide a safe and controlled environment for surgeons to practice and rehearse surgical procedures (Liu, Tendick, Cleary, & Kaufmann, 2003).

Despite the potential for HMDs to enhance perception and performance in complex work environments, HMDs are confronted by many technical and ergonomic challenges, including optical distortion, suboptimal resolution, FOV limitations, time delays, and helmet fit and discomfort. While these technical and ergonomic limitations have been shown to adversely affect performance and operator workload, HMDs can also cause simulator sickness

(Stanney, Hale, Nahmens, & Kennedy, 2003; Stanney et al., 1998). Simulator sickness is a significant problem in synthetic environments because in some cases the symptoms are severe enough for users to discontinue use (Stanney et al., 2003; Stanney, Lanham, Kennedy, & Breaux, 1999) and for some users the symptoms may linger for a period of time after use, potentially compromising operator safety and acceptance (Stanney & Kennedy, 1998; Stanney, Kingdon, & Kennedy, 2002).

Although there are problems inherent with the use of HMDs, they may serve as a promising solution to the problem of display occlusion in ABM work domains. The utility of HMDs in multi-task environments remains uncertain therefore it is important to identify operationally relevant task environments for which HMDs are best suited. The purpose of the present investigation was to evaluate various display technologies for reducing the effects of occlusion on task performance in ABM work domains during simulated air-battle scenarios.

Method

Participants

Six males and six females between the ages of 18 and 34 (M = 23.83) participated in the experiment. All participants reported normal or corrected-to-normal vision in both eyes. Individuals were paid for their participation.

Apparatus

The study was conducted in a medium-fidelity simulated AWACS environment. A stereo headset was required to hear audio tones and radio calls. A calculator was provided for use by the participants for time and distance calculations.

Two commercial-off-the–shelf HMDs were evaluated during the experiment. A monocular HMD, the MicroOptical Instrument Viewer (SV-9), was tested (MicroOptics). This HMD was a VGA clip-on (glasses) display with a color LCD that presented a full-size image right in front of the eye (either right or left eye). It provided a 20 degree field of view with a resolution of 640 × 480. In addition, the Sony Personal LCD binocular HMD was tested (Glasstron). This HMD was a small, lightweight (5.3 oz) VGA head-wearable display with two 1.55 million dot LCDs and a resolution of 640 × 480. It provided a television viewing experience comparable to watching a 30-inch screen from a distance of approximately 4 feet.

Primary Task

Participants were asked to control an air battle involving the re-targeting of strike aircraft. The participants were required to perform distance measurements and calculations to determine if strike aircraft could be re-directed to various targets and/or an air refueler using information provided on a re-targeting form. Participants were required to look up strike aircraft call signs, preplanned air refuelers, and planned refueling times on the re-targeting form. They also needed to determine distances using the on-screen measuring features. Worksheets and a calculator were provided for use by the participants.

Secondary Tasks

At random times throughout each mission, 4 radio frequency calls occurred (2 via audio and 2 via a displayed text message) requiring the participants to look up and enter a new radio frequency from a form.

During each trial, participants also received 4 authentication tasks requiring the participants to search for and enter an authentication code found on an authentication form.

The presentation of the 3 forms (re-tasking, radio frequency change, authentication) for each mission was accomplished using one of the possible display technologies (paper forms, forms displayed electronically on the monitor, forms displayed with the monocular HMD and forms displayed with the binocular HMD). The three forms were available in Excel format.

All participants received a training protocol that was divided into three functional areas: 1) operator workstation and tactical display control; 2) measuring and calculation training without the secondary tasks; and 3) measuring and calculation training with the secondary tasks.

During training it was explained to the participants that all of the tasks were important and to complete them quickly and accurately. The participants were instructed to develop strategies to aid them in the completion of all the tasks in the allotted 10 minute mission time.

Experimental Design

A within-subjects design was employed. Display technology (paper forms, forms displayed electronically on the monitor, forms displayed with the monocular HMD, and forms displayed with the binocular HMD) was the manipulated variable**.** Each participant completed 3 missions (trials) for each display technology for a total of 12 trials. The maximum duration of each mission was 10 minutes. After completion of each mission, the participants were asked to rate their subjective impressions of mental workload and situational awareness (SA). The entire experiment, including training, lasted approximately 5-6 hrs for each participant.

Subjective Measures

The NASA-TLX (Hart & Staveland, 1988) sub-scales were used for ratings of mental workload. The Measures of Situation Awareness (3-D SART) questionnaire (Taylor, 1990), with an additional question asking the participant to rate their overall SA, was also administered. Both scales were rated by the participants following each mission.

Results

Primary Task Performance

The data collected during the trials was analyzed using a 4 (display technology) factor Analysis of Variance (ANOVA) for both percentage of correct responses and the response times. The results of the ANOVA conducted for the primary task indicated that there was a significant difference in the percent of correctly re-tasked strike aircraft, $F(3, 33) = 5.25$, $p < .01$. This analysis, depicted in Figure 1, indicates that participants responded correctly more often when the information about the re-tasking was available via paper. This was followed by the screen condition, the Glasstron HMD condition, and finally the MicroOptics condition. The response times for this task failed to reach a significant difference for the four display technologies available.

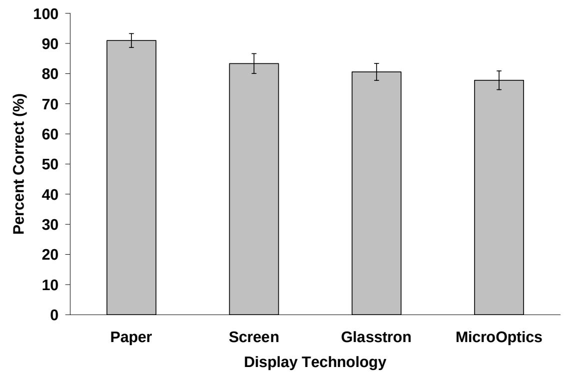


Figure 1. *Percent of correct responses for the re-tasking evaluation as a function of display technology.*

Secondary Task Performance

A similar statistical strategy was employed for the percent correct and response times for the secondary tasks. The results indicated that there was a significant difference for the percent of correct responses for the radio frequency change task, $F(3, 33) = 19.30$, $p < .01$. This result, depicted in Figure 2, suggests that participants responded correctly in a similar manner to that of the primary task; more often when the information was available via paper, followed by the screen, the Glasstron HMD, and finally the MicroOtics HMD.

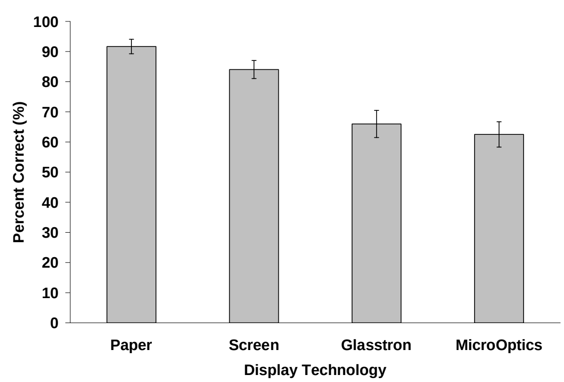


Figure 2. *Percent of correct responses for the radio frequency change task as a function of display technology.*

The results of the ANOVA conducted on the response times for the radio frequency change task was also significantly different for the different display technologies available, $F(3, 33) = 23.53$, $p < .01$. The response times for this task, illustrated in Figure 3, are inversely related to the percentage of correct responses. That is, the response times for the radio frequency task were fastest when that information was available in the paper condition. Response times lengthened in the screen condition, followed by the Glasstron HMD and the MicroOptics HMD.

The ANOVA conducted for the authentication task revealed that there was not a significant difference for the percentage of correct responses but the response times did espouse a significant difference for this task, $F(3, 33) = 41.28$, $p < .01$. This difference is virtually the same as the result found for the radio frequency change task. The shortest response times were those that were obtained when the participants had the information available via paper (M = 18.38s, SE = 0.87s), followed by the screen condition (M = 23.41s, SE = 0.91s), the Glasstron HMD (M = 33.66s, SE = 1.75s), and the MicroOptics HMD (M = 34.40s, SE = 1.56s).

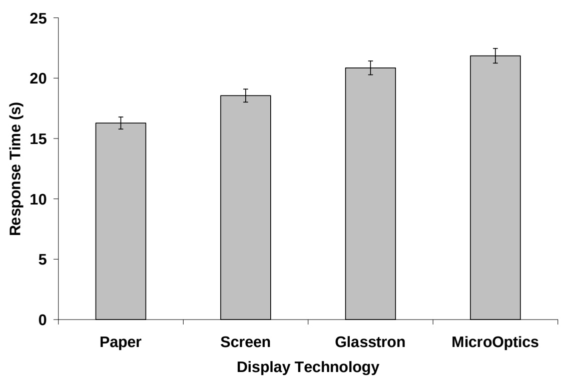


Figure 3. *Response times for the radio frequency task as a function of display technology.*

Subjective Measures

The NASA-TLX sub-scale scores were averaged to yield one workload score for each trial. This score was used in an ANOVA analogous to that described previously. The results indicated that there was a significant effect for the display technology on the workload ratings, $F(3,33) = 5.72$, $p < .01$. This effect, illustrated in Figure 4, indicates that participants rated their workload highest while using the MicroOptics HMD, followed by the Glasstron HMD, the screen condition, and lowest when the information was available via paper. Participant ratings of Situation Awareness failed to differ significantly for the display technology utilized.

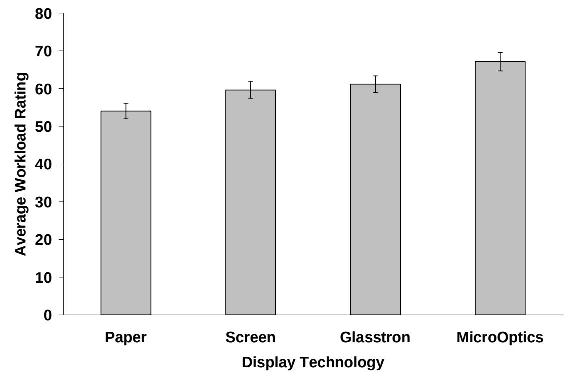


Figure 4. *Average participant workload ratings as a function of display technology.*

Discussion

This is the first in a series of studies examining the utility of HMDs in an Air Battle Management environment. The results indicate that the HMDs selected did not produce a significant performance benefit. Further, the workload reported by the participants suggests that they experienced the lowest workload when the information was available via the paper medium. Upon examination of the results, it was posited that the reason for the lack of a performance benefit may be due to the nature of the tasks the participants were required to perform. Namely, these results may be significantly influenced by the lack of complexity in the required tasks. All of the forms that were used in the information retrieval were one page or less in length. It was suggested that this one page length may not be representative of the types of tasks that may be amenable to the utilization of HMDs. Further, it was suggested that the information operators typically need to access is often found in sources that are comprised of several, if not several hundred pages. The next experiment in this series will utilize more complex tasks to examine the potential benefit HMDs may provide.

References

Conn, E.S. (2003). *JAVA-BASED performance oriented visualization system* (Tech. Rep. No. AFRL-IF-RS-TR-2003-87). Wright-Patterson Air Force Base, OH: Air Force Research Laboratory.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: Elsevier Science/North Holland.

Liu, A., Tendick, F., Cleary, K., & Kaufmann, C. (2003). A survey of surgical simulation: Applications, technology, and education. *Presence, 12,* 599–614.

ModSAF 5.0, Advanced Distributed Simulation Technology (1999). *ADST-II-CDRL-MODSAF5.0-9800324*.

St. John, M., Manes, D. I., Oonk, H. M., & Ko, H. (1999). Workspace control diagrams and head-mounted displays as alternatives to multiple monitors in information-rich environments. *Proceeding from the Human Factors and Ergonomics Society 43rd Annual Meeting,* 438-442.

Stanney, K. M., Hale, K. S., Nahmens, I., & Kennedy, R. S. (2003). What to expect from immersive virtual environment exposure: Influences of gender, body mass index, and past experience. *Human Factors, 45,* 504-520.

Stanney, K. M. & Kennedy, R. S. (1998). Aftereffects from virtual environment exposure: How long do they last? *Proceeding from the Human Factors and Ergonomics Society 42nd Annual Meeting*, 1476-1480.

Stanney, K. M., Kingdon, K. S., & Kennedy, R. S. (2002). Dropouts and aftereffects: Examining general accessibility to virtual environment technology. *Proceeding from the Human Factors and Ergonomics Society 46th Annual Meeting*, 2114-2118.

Stanney, K. M., Lanham, D. S., Kennedy, R. S., & Breaux, R. (1999). Virtual environment exposure drop-out thresholds. *Proceeding from the Human Factors and Ergonomics Society 43rd Annual Meeting*, 1223-1227.

Stanney, K., Salvendy, G., Deisinger, J., DiZio, P., Ellis, S., Ellison, J., Fogleman, G., Gallimore, J., Singer, M., Hettinger, L, Kennedy, R., Lackner, J., Lawson, B., Maida, J., Mead, A., Mon-Williams, M., Newman, D., Piantanida, T., Reeves, L., Riedel, O., Stoffregen, T., Wann, J., Welch, R., Wilson, J., & Witmer, B. (1998). Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda (Report sponsored by the Life Sciences Division at NASA Headquarters). *International Journal of Human-Computer Interaction, 10,* 135-187.

Stelle, J. A., Reynolds, B. S., Rash, C. E., Peterson, R. D., & Leduc, P. A. (2003). The role of the pilot's night vision system (pnvs) and the integrated helmet and display sighting system (ihadss) in ah-64 apache accidents. *Proceedings of Helmet- and Head-Mounted Displays VIII: Technologies and Applications, SPIE Vol. 5079,* 31-40.

Taylor, R.M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In Situational Awareness in Aerospace Operations (AGARD-CP-478, pp. 3-1 and 3-17). Neuilly Sur Seine, France: NATO-AGARD.

DEVELOPMENT AND VALIDATION OF A SURVEY TO ASSESS SAFETY CULTURE IN AIRLINE MAINTENANCE OPERATIONS[1]

Alyssa Mitchell Gibbons
Department of Psychology – University of Illinois at Urbana-Champaign
Champaign, Illinois

Terry L. von Thaden
Aviation Human Factors Division – University of Illinois at Urbana-Champaign
Savoy, Illinois

Douglas A. Wiegmann
Aviation Human Factors Division – University of Illinois at Urbana-Champaign
Savoy, Illinois

This paper describes the development and validation of a survey to assess safety culture in airline maintenance operations according to the five-factor model of safety culture proposed by Wiegmann et al. (2002). Maintenance technicians at two FAR Part 121 scheduled passenger airlines (N = 109 and 76) completed the original version of the survey. The results yielded useful diagnostic information about the safety culture of each airline, but factor analyses indicated that the five-factor model may not be adequate to describe the data. A more complex model is proposed and modifications to the survey are suggested.

Safety Culture

Aviation organizations are becoming highly interested in understanding safety culture and how it can be improved. Safety culture can be defined as "the enduring value and priority placed on worker and public safety by everyone in every group at every level of an organization" (Wiegmann, Zhang, von Thaden, Sharma, & Mitchell, 2002). Wiegmann et al. (2002) reviewed the safety culture literature across a number of industries and identified five critical indicators of an organization's safety culture:

Organizational Commitment (OC): the organization's commitment to safety, as expressed by upper management;
Managerial Involvement (MI): the active involvement of mid-level managers or supervisors in promoting safety;
Employee Empowerment (EE): the degree to which individual employees are empowered to make safety a priority;
Accountability System (AS): the system by which employees are held accountable for acting unsafely; and
Reporting System (RS): the quality and usability of the system for reporting and processing safety information.

While strength in one area can compensate to some extent for deficiency in another (e.g., strong employee empowerment may limit the negative impact of poor management involvement), strength in all areas is the hallmark of a culture that truly promotes safety to the fullest.

Commercial Aviation Safety Survey

The Commercial Aviation Safety Survey (CASS) was developed, based on the five-factor model described above, to assist airlines in diagnosing strengths and weaknesses within their safety cultures so that the weaknesses can be addressed. The flight operations version of the CASS was created first, with items based on existing safety culture inventories from a number of industries. Wiegmann, von Thaden, Mitchell, Sharma, & Zhang (2003) provide a detailed description of the development of the flight operations survey. The development of the maintenance survey is the focus of the present report. The maintenance version of the survey is designed to reflect the same structure as the flight operations version (that is, the same five indicators of safety culture), but to use terminology and describe behavior appropriate to the maintenance function.

Several steps were taken to develop the survey in such a way that it paralleled the structure of the flight operations survey but contained items directly relevant to maintenance professionals. The flight operations survey contained 89 items. Thirty-eight of

[1] This research was supported by Federal Aviation Administration Award No. DTFA 01-G-015. Kathy Fazen was the technical monitor. The views expressed are those of the authors and do not necessarily reflect those of the Federal Aviation Administration. The authors thank William Geibel and Marcus Credé for their assistance in the survey development process.

these were judged as applicable for maintenance with minimal revisions. An additional 13 items were modified more extensively to reflect the intent of each original flight operations item in a maintenance context. For example, the item “management expects pilots to push the weather” was replaced by the item “supervisors never pressure inspectors to sign-off on borderline work.” Maintenance technicians are not at all likely to encounter the scenario described in the first item, but both items represent the same basic idea: a specific, common situation in which the responsible manager pressures a subordinate to behave in an unsafe manner.

At the time that the maintenance survey was being developed, one preliminary test of the flight operations survey had already been conducted, so general feedback from that test was incorporated into the revisions. Apparently confusing or ambiguous items were excluded, as were items that did not appear to have clear parallels in maintenance operations. Additional items were chosen and/or written to fill in the gaps left by the excluded items. The original safety culture inventories used to create the flight operations survey were consulted again, to see whether we had overlooked any items appropriate for maintenance. This search yielded six items. Twenty new items were written based on the extensive airline maintenance experience of one of the researchers, who pointed out situations and attitudes common in such an environment.

The final version of the maintenance survey contained 84 items. As in the flight operations survey, respondents were instructed to use a 7-point Likert-type response scale to indicate their agreement or disagreement with each item. A rating of 1 indicated that the respondent “strongly disagree(d)” with the item and a rating of 7 indicated that he or she “strongly agree(d).” The center point of the rating scale, 4, was labeled “neither agree nor disagree.” Space was provided beside each item for respondents to write comments if they chose.

Initial Results from Two Airlines

Maintenance personnel from two FAR Part 121 passenger airlines completed the survey. Participants returned surveys directly to the researchers. They were assured that their responses would remain confidential and they were not asked to provide their names or other personally identifying information. No compensation was offered to participants or their organizations.

A total of 1148 surveys were distributed: 860 to employees of Airline A and 288 to Airline B. One hundred and nine of the Airline A surveys and 76 of the Airline B surveys were returned, for response rates of 13% and 26% respectively. At Airline A, most respondents (74%) described their primary job responsibility as “Aircraft Technician;” at Airline B, respondents were more evenly divided between technicians (40%) and supervisory positions (Line Manager, Lead Technician, Inspector, or Manager; 51% combined).

Dimension Scores. Scores for each airline were calculated for each of the five dimensions of safety culture as the mean of participants’ responses to the items in each dimension scale. Items indicating a negative safety culture (e.g., “My airline is more concerned with making money than being safe.”) were reverse coded. All five scales showed acceptable levels of reliability for both airlines (α = .74 – .94). Dimension scores for both airlines appear in Figure 1.

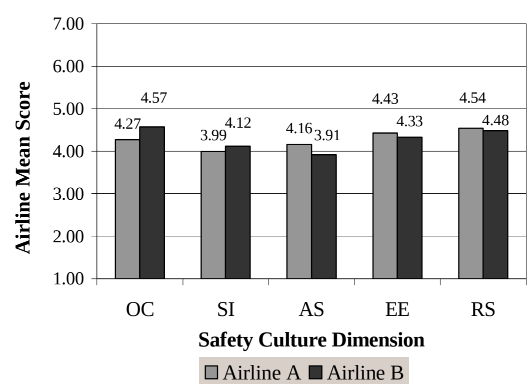


Figure 1. *CASS scores for two FAR Part 121 airlines.*

Both of these airlines appear to have “middle-of-the-road” safety cultures, with scores near the neutral point (4.0) in most areas. However, the pattern of the dimension scores suggests different areas of strength and weakness for each airline, implying that the actual safety cultures experienced by employees are quite different. Reporting systems are strong at both airlines; they are the strongest area at Airline A, while organizational commitment is Airline B’s strongest dimension. Airline A needs improvement in supervisory involvement, while Airline B needs to improve its accountability system. Analyses of individual item responses and respondent comments further supported these overall impressions. All scales were negatively correlated with technicians’ perceptions of risk at both airlines. The maintenance CASS appears to be a useful diagnostic tool. The items can be grouped together into reliable scales to provide a broad-level picture of the organization or analyzed individually to identify specific strengths and weaknesses, providing useful information to airline management seeking to improve safety culture.

Factor Analysis of the Maintenance CASS

Analytical Strategy

To validate the five-dimensional model of safety culture proposed above, we conducted confirmatory factor analyses (CFAs) using the Mx software package (Neale, 2002). We conducted an overall CFA for the five factor model and then tested single-factor models for each of the five dimension scales individually. In all analyses, model fit was tested by considering the overall chi-squared value for the model (X^2), the root mean squared error of approximation (RMSEA), the normed fit index (NFI), the Tucker-Lewis index (TLI) and the relative noncentrality index (RNI; see Neale, Boker, Xie, & Maes, 2004 for definitions and citations for all fit indices). Models are usually considered to fit well when the X^2 value is nonsignificant compared to the degrees of freedom, the RMSEA is below .10, and the NFI, TLI, and RNI are above .90 (McDonald & Ho, 2002). Given the relatively small sample used in this study, we considered a model to fit well when most of these criteria were met.

If a model did not fit well, we considered the matrix of residual discrepancies between the observed correlation matrix and that expected under the model. When an item showed large residual correlations (> .15) with other items, we considered whether the item might have been confusing or ambiguous, whether it was highly correlated with only one other item (introducing instability into the model), or whether it showed a pattern of large residuals with other items that might suggest the existence of another factor. In the first two cases, the item was discarded and the fit of the model without that item was assessed. In the third case, items were grouped logically into subfactors and the fit of the new multifactor model was assessed. Improvement in fit was assessed by means of chi-squared difference tests, which compare the fit of the original model with the fit of the revised model. A significant difference implies that the revised model fits significantly better than did the original model. If the large residuals within a scale could not be resolved through these methods, exploratory factor analyses (EFAs) were conducted using the COFA (see McDonald, 1999) software program to determine whether multiple factors were needed to describe the data. Exploratory factor analyses were subsequently retested with confirmatory factor analyses so that the fit obtained could be compared to that of the original models.

The data from Airline A were used in the initial analyses, because the small sample size from Airline B was not sufficient to allow a test of the full model. The data from Airline B were used as a cross-validation sample for the revised versions of the individual factor scales.

Overall Model Fit. The first model tested was the one hypothesized: a five-factor solution with each item loading on the dimension it was intended to measure. This model fit the data poorly: $X^2_{2765} = 5660.27$, $p < .01$; RMSEA = .11; NFI = .30; TLI = .43; RNI = .45. Further, 10% of the residual correlations had absolute values greater than .15. One possible explanation for the poor fit of the five-factor model is that respondents did not discriminate between dimensions of safety culture when completing the survey, but rather based their responses on their overall perception of the safety culture as good or bad. If this were the case, a single-factor model in which all items simply reflect the overall positive or negative safety culture of the organization would fit well. The single-factor model also fit poorly, however: $X^2_{2774} = 5711.40$, p<.01; RMSEA = .11. In fact, the fit of the single-factor model was significantly poorer than that for the five-factor model, $\Delta X^2_9 = 51.14$, $p < .01$.As the data could not be described by either the five-factor model or a single general factor, the dimension scales were analyzed individually to identify specific sources of misfit.

Accountability System. The single-factor model for the accountability system scale showed acceptable fit: $X^2_{27} = 32.93$, p=.20; RMSEA = .05, NFI = .84, TLI = .95, RNI = .96. Examination of the residuals suggested that two pairs of items shared particularly high correlations, implying that the items in each pair may measure the same thing to such an extent as to be redundant. However, removing one item from each pair did not significantly improve the fit of the model, so the items were retained. Cross-validation with data from Airline B showed reasonably acceptable fit, $X^2_{27} = 48.75$, p=.01; RMSEA = .11; NFI = .87; TLI = .91; RNI = .94.

Reporting System. The single factor model showed acceptable fit for the reporting system scale, $X^2_{35} = 42.63$, $p = .18$; RMSEA = .05; NFI = .85; TLI = .96; RNI = .97. Low correlations between two pairs of items resulted in high residuals (> .15) for those pairs, but as all four items had high correlations with the other items in the scale and the overall fit of the model was good, they were retained. However, in the cross-validation sample from Airline B, the single-factor model did not fit as well, $X^2_{35} = 70.13$, $p = .00$; RMSEA = .12; NFI = .75; TLI = .81; RNI = .85.

Removing the item "I am familiar with the system for formally reporting safety issues in my airline" improved the fit of the model in the Airline B sample, $X^2_{27} = 49.86$, $p = .01$; RMSEA = .11; $\Delta X^2_8 = 20.26$, $p<.01$; NFI = .80; TLI = .86; RNI = .89; and removing it from the Airline A data improved the fit slightly, but not significantly $X^2_{27} = 33.54$, $p = .18$; RMSEA = .05; $\Delta X^2_8 = 9.10$, $p = .33$. In light of that evidence, the item was retained.

Supervisory Involvement. The initial single factor model did not quite fit the supervisory involvement scale well, $X^2_{77} = 108.78$, $p = .01$; RMSEA = .07; NFI = .83; TLI = .93; RNI = .94. Examination of the residual matrix indicated that a large number of the discrepancies were related to two items. Respondent comments on one of the items indicated that the item was interpreted differently by different respondents, but the reason for the misfit of the other item was unclear. Removing both items, however, improved the fit of the model, so that the model containing twelve items fit acceptably, $X^2_{54} = 63.59$, $p = .17$; RMSEA = .04; $\Delta X^2_{11} = 24.11$, $p = .01$; NFI = .85; TLI = .95; RNI = .96. In the Airline B sample, the fit of the twelve-item model was similar, but not quite so good, $X^2_{54} = 84.73$, $p < .01$; RMSEA = .09; NFI = .80; TLI = .90; RNI = .92.

Employee Empowerment. The single factor model for the employee empowerment scale did not fit particularly well, $X^2_{54} = 87.42$, $p<.01$; RMSEA = .08, NFI = .73, TLI = .84, RNI = .87. An attempt to separate the items into two factors (technicians' authority to improve safety and their safety professionalism) based on large residuals and logical relationships among items yielded only slightly improved fit ($X^2_{53} = 76.26$, $p = .02$; RMSEA = .07) and a further division into three factors (authority, professionalism, and peer influence) did not fit better ($X^2_{51} = 75.50$, $p = .01$; RMSEA = .07). Consequently, an exploratory factor analysis (EFA) was conducted to investigate the structure of the scale. A three-factor model was tested first, because the three-factor model suggested above showed (though barely) the best fit of the three. The promax rotated solution identified three factors that were in many ways similar to the three factors suggested by conceptual grouping. The first factor appears to reflect supervisors' respect for technicians in safety matters (authority), the second describes technicians' personal pride in upholding safety standards (professionalism), and the third indicates a peer culture that supports safety (peer influence). In a CFA, this model showed acceptable fit, $X^2_{51} = 65.56$, $p = .08$; RMSEA = .05. One conceptual difficulty remained in that two items, "Everyone routinely performs the operational checks after the work is completed," and "Everyone routinely re-inspects each other's work or has someone inspect their work before return to service," were clearly similar in content, but loaded on different factors. However, the former item had near-equal loadings (.26 and .25, respectively) on both the professionalism and peer culture factors. Moving this item to the peer culture factor actually slightly improved the fit of the model, $X^2_{51} = 61.14$, $p = .16$; RMSEA = .05. In the interest of parsimony, a second exploratory analysis was conducted requesting only two factors. The factors identified by the promax rotation were identical to those suggested by the conceptual two-factor grouping. As that model had already been shown to fit poorly, the three-factor model for the employee empowerment dimension was retained.

The three-factor model appeared to fit the cross-validation data from Airline B well, $X^2_{51} = 45.54$, $p=.69$; RMSEA = .00. However, some of the fit indices were inappropriately high: NFI = .75; TLI = 1.06; RNI = 1.05. This suggests empirical underidentification, a condition that occurs when the observed correlations between variables in a sample are near zero. This is most likely a function of the small Airline B sample size, but it prevents us from being able to draw conclusions about the cross-sample validity of the three-factor employee empowerment model.

Organizational Commitment. The single-factor model did not fit the data well for the organizational commitment scale, $X^2_{434} = 732.23$, $p < .01$; RMSEA = .08; NFI = .55; TLI = .73; RNI = .75. Of the 465 residual correlations, ninety-nine were greater than .10, with 34 greater than .15. This suggests that a multi-factor model is necessary to describe the items in this scale – such pervasive residuals are not likely to be resolved by removing a few items. To identify a starting point for conceptually grouping these items, we looked to the parallel analysis that had previously been conducted for the flight operations survey. That investigation used an exploratory factor analysis to identify three factors: upper management attitude toward safety, use of preventive safety practices (such as safety training), and commitment of organizational resources to safety. The maintenance items were correspondingly grouped into similar factors and a three-factor model was tested. However, that model showed only small (but significant) improvement in fit over the single-factor model, $X^2_{431} = 704.48$, $p = .00$; RMSEA = .08.

A series of EFAs was then conducted using COFA. Two-, three-, four-, and five-factor solutions were tested, but the four-factor model showed the best fit in subsequent CFAs: $X^2_{399} = 550.69$, $p = .00$; RMSEA

= .06, with 21 residuals greater than .15. Fit indices for this model approached acceptable levels, NFI = .69, TLI = .88, RNI = .89. The first three factors in this model appeared to represent management attitude (e.g., "Unsafe behavior is not tolerated in my company"), allocation of resources (e.g., "Tool control, calibration, and equipment certification are closely monitored by my company"), and quality of safety training. .The fourth factor contained only three items, and no conceptual relationship between these was readily apparent, except perhaps that all were rather indirect assessments of safety culture. As this factor was not clearly interpretable and may have simply consisted of poor items, another confirmatory analysis was conducted excluding those three items (and thus the fourth factor). This new three-factor model (consisting of attitude, resource, and training factors) did not yield a significant improvement in fit over the four-factor model (X^2_{321} = 475.63, p = .00; RMSEA = .07; ΔX^2_{78} = 75.05, p=.57). However, in the revised model, it became apparent that many of the large residuals were associated with one item. Excluding this item from the new three-factor model resulted in a significant improvement in fit for that model (X^2_{296} = 421.30, p = .00; RMSEA = .07; ΔX^2_{25} = 54.33, p<.01), and the resulting model also fit significantly better than the four-factor model (ΔX^2_{103} = 129.38, p=.04). Fit indices for this model were similar to those for the four-factor model, NFI = .72, TLI = .88, RNI = .89. Eight large (> .15) residuals remained, but no item was connected with more than one of these, and no logical connections between pairs of items sharing large residuals were apparent. The revised three-factor model (attitude, resources, and training) was retained.

Again, data from Airline B were used to cross-validate the revised model. The three-factor model did not fit particularly well, X^2_{296} = 480.08, p = .00; RMSEA = .10, NFI = .61; TLI = .77; RNI = .79. This suggests that these factors should be used with caution in future research, as they may reflect idiosyncratic characteristics of Airline A rather than the general structure of organizational commitment across airlines.

Revised Model Overall Fit. When all revisions are taken into account, the new model contains a total of nine factors (the organizational commitment and employee empowerment scales were each divided into three factors). The original survey contained seventy-six items, but seven items were deleted in the revision process, so the revised model contained only sixty-nine. A confirmatory factor analysis was conducted to test the fit of the revised model. Again, however, the full model did not fit particularly well, X^2_{2246} = 4099.49, p < .01; RMSEA = .10; NFI = .38; TLI = .54; RNI = .56. While these values represent an improvement in fit compared to the original model, they still fall short of acceptable levels. Of the 2415 residuals, 231 (9.6%) had absolute values above .15.

To determine whether the observed improvement in fit was due to the regrouping of items or merely to the elimination of poor items, an additional confirmatory factor analysis was conducted using only the sixty-nine items in the revised model but grouping them into the original five factors. Again, this model fit poorly overall, X^2_{2267} = 4236.75, p < .01; RMSEA = .10. The fit of the revised (nine-factor) model was significantly better than that of this five-factor model, ΔX^2_{21} = 137.26, p < .01, but comparison of the other fit indices (NFI = .36; TLI = .52; RNI = .54.) suggests that the difference is slight.

Conclusions

While the results of the factor analysis generally supported the three of the five factors, the organizational commitment and employee empowerment factors remained problematic, and even the revised complete model did not show a good fit to the data. These findings are consistent, however, with the corresponding analysis of the flight operations survey. That survey also indicated a need to divide organizational commitment and employee empowerment into three subfactors each, and the subfactors identified in that analysis correspond conceptually in many ways to the subfactors identified here. The correspondence is not exact, but these findings do suggest two conclusions: (1) While the accountability system, reporting system, and management/supervisory involvement scales appear to represent well-defined, unitary constructs, the pilot/employee empowerment and organizational commitment scales represent more complex phenomena that require further consideration; and (2) within those two scales, several consistent themes emerge that provide insight into how those constructs might be better defined in future.

Specifically, the concept of employee (or pilot) empowerment seems to encompass several distinct elements: the authority granted to employees by the organization, the authority and personal responsibility assumed by employees, and the positive or negative impacts of the peer culture regarding safety. With respect to organizational commitment, respondents in both flight operations and maintenance appeared to distinguish between the "talk" (statements and policies) and the "walk" (actions and commitment of resources) of their organizations.

The analysis of the flight operations survey concluded with a conceptual revision of the scale, based on input from the factor analysis and from respondent comments. For example, the pilot empowerment subfactors were linked with the accountability system factor as aspects of an "Informal Safety System" second-order factor and the reporting system was similarly divided to indicate different parts of the reporting process. Given the strong conceptual similarity between the flight operations analysis and the results reported here, we considered whether a similar structure could be adopted for the maintenance survey. Again, we considered respondents' comments as well as the factor analyses to identify problematic items or areas of concern to technicians that might have been overlooked in the original survey. The revised model for maintenance appears in Table 1.

Table 1. *Revised model of safety culture maintenance.*

Construct	Factors	Subfactors
Overall Safety Culture	Organizational Commitment	Safety Values Safety Fundamentals Work Environment Safety Training
	Supervisors	Supervisory Involvement Maintaining Standards
	Informal Safety System	Accountability Technicians' Authority Professionalism
	Formal Safety System	Reporting System Response & Feedback Safety Personnel

The informal safety system and reporting system factors from the revised flight operations survey were retained for the revised maintenance version. For the organizational commitment factor, the safety values and safety fundamentals subfactors from the flight operations survey were kept, but subfactors for safety training and a safe work environment were added. The supervisory involvement factor was retitled "Supervisors" and contained only two subfactors: supervisory involvement and "maintaining standards". This latter subfactor referred to supervisors' consistent enforcement of high safety standards. This reorganization required the creation of several new, specific items to ensure that each subfactor had enough items to be stable in future analyses. Items were also excluded if they seemed less relevant than or redundant with other items in the same scale. We also revised item wording wherever it appeared that an item might have been ambiguous or confusing.

While this initial test of the maintenance version of the CASS did not provide solid support for the five-factor model of safety culture, it nevertheless provided useful information on which further revisions can be based. The five scales proved useful as a diagnostic tool for identifying strengths and weaknesses of two airlines' safety cultures. Detailed factor analyses indicated that the accountability system, reporting system, and supervisory involvement factors represented fairly unitary constructs, which the employee empowerment and organizational commitment factors were more complex. As this is consistent with the findings of the flight operations survey, it seems likely that this reflects true complexity in the construct rather than only measurement error. When combined with respondents' substantive comments on the items, the factor analyses yielded information that was useful in creating a revised model of maintenance safety culture parallel to that created for flight operations . This new model formed the basis for an extensive revision to the maintenance CASS that may be tested in future research.

References

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Laurence Erlbaum.

McDonald, R. P., & Ho, M-H. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*(1), 64-82.

Neale, M. C. (2002). *Mx*. Richmond: Virginia Commonwealth University.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2004). Mx: Statistical Modeling. Richmond: Virginia Commonwealth University.

Wiegmann, D.A., von Thaden, T. L., Mitchell, A. A., Sharma, G., & Zhang, H., (2003). Development and initial validation of a safety culture survey for commercial aviation. Technical Report AHFD-03-3/FAA-03-1. Prepared for the Federal Aviation Administration, contract DTFA 01-G-015.

Wiegmann, D., Zhang, H., von Thaden, T., Sharma, G., & Mitchell, A. (2002). A synthesis of safety culture and safety climate research. Technical Report ARL-02-3/ FAA-02-2. Prepared for the Federal Aviation Administration, contract DTFA 01-G-015.

A CONCEPTUAL FRAMEWORK FOR STUDYING SAFETY CLIMATE AND CULTURE OF COMMERCIAL AIRLINES

Dr. Sharon Glazer
Amber R. Laurel
Rangapriya Kanan Narasimhan
San Jose State University
San Jose, California

A comprehensive safety climate and safety culture framework, which can be utilized to assess various predictors and consequences of safety climate and to assess airline's safety culture in relation to one another, is presented. The framework depicts a process whereby individual, group, and organizational predictor variables, through perceived safety climate, affect first level outcomes. First level outcomes can lead to direct costs for the organization, as well as lowered productivity. In the framework, individual and environment variables are purported to moderate the relationship between work-related events and safety climate. Motivation is also expected to mediate the relationships between predictors and safety climate, as well as predictors and individual level consequences. Overall, organizational culture and environment are likely to affect safety climate and safety culture.

Introduction

To date, there is a lack of comprehensive and coherent safety culture or safety climate frameworks (Mearns & Flin, 2001). Studying safety climate and culture of airlines is a difficult undertaking; therefore it is often the case that problems are solved reactively and the focus is on mechanics of mishap(s). Examining safety climate and culture from an organizational psychological perspective, however, could provide a more holistic understanding of why and how mishaps occur, and provide a predictive model for preventing them. The basis of an organizational psychological perspective is people's perceptions of organizational processes (e.g., structure, selection, reward policies), which are often the root of mishaps (Reason, 1997). Therefore, in this paper we present a comprehensive safety climate and safety culture framework (see Figure 1), which can be utilized to assess various predictors and consequences of safety climate and to assess airline's safety culture in relation to one another.

Conceptual Framework

Our framework is based on Lazarus and Folkman's (1984) transactional model for studying occupational stress. The transactional model demonstrates that process variables mediate the relationship between predictors and outcomes. Safety climate is depicted in our model as a process variable; it's through safety climate that predictors will affect outcomes. It provides context for why certain consequences occur due to work-related events.

Predictors of Safety Climate/Culture

Individual Predictors. Individual predictors are variables that reflect characteristics of the people who are employed in an organization and the characteristics of the jobs in which they work. Two individual predictors identified in our framework are job characteristics and personal characteristics. *Job characteristics* describe attributes of a job, such as task involvement, job autonomy and responsibility, skill discretion, physical demands, work hours, shift patterns, and fatigue. Previous research has found that organizational members contribute more in ensuring safe operations when provided autonomy and responsibility within their work tasks (Parkes & Bochner, 2001), as responsibility can lead to a sense of pride in maintaining a good safety record (von Thaden et al., 2003). Additional characteristics of one's work environment include *work schedules, work hours, shift patterns, and fatigue.* Research has shown that, demanding pilot schedules leads to fatigue and subsequent performance problems and errors (Bourges-Bougrine et al., 1999). Finally, *physical demands* reported by flight crew members, such as inadequate cockpit design and experience of fluctuations between hot and cold temperatures, noise, altitude pressure, and acceleration (e.g. Orlady & Orlady, 1999) can have detrimental effects on employees' health, and subsequently, safe flight operations (Gadd, 2002).*Personal characteristics*, such as s*afety consciousness*, are associated with taking safety precautions, and low levels of safety consciousness can lead to adverse outcomes, such as accidents (Behn et al., 1999). *Safety competence* (Gadd, 2002) has been shown to increase likelihood of safe flight operations (Hofmann et al., 1995).

Group Predictors. Group level predictors are classified into two subcategories: *leadership* and *psychosocial stressors*. Previous research has found that *leadership* affects the way subordinates perceive safety (e.g. Zohar, 2002) and lack of strong leadership was directly related to incidents and accidents amongst aircrews during simulator training exercises (Kanki, 1996). Safety climate was found to be affected by *management commitment* (e.g. Wiegmann et al., 2002). Mearns et al. (2001) found that employees' perceptions of managements' commitment to safety was positively correlated to satisfaction with safety actions. Other leadership aspects that can affect safety climate are *task orientation and goal setting (*Tuttle et al., 1975), as well as *innovation or risk behaviors*. Leaders are also pivotal in *monitoring safe practices (*Huettig et al., 1999) which is central to pilot decision making, and consequent flight safety. *Psychosocial stressors* consist of variables related to role behaviors and perceptions, which entail role conflict, role overload, role ambiguity, interpersonal relationships, and communication. Stressors can have human and financial costs (e.g., turnover, poor work performance, accidents, and fatalities; Tuttle et al., 1975). Role overload (i.e., performance pressure) has been found to be a strong predictor of injury (e.g. Zohar, 2000) and can lead to avoidance coping methods (Dillenger et al., 2003). Avoidance coping can adversely affect accident prevention, e.g., behavioral disengagement was chosen as the first choice of coping strategy amongst student pilots (Dillenger et al.). Also, C*ommunication* of safety-related information must occur upward, as well as downward, and must be accessible to anyone needing it to perform well. In an aviation context, when pilots do not engage in positive briefings with the other crewmembers, they can be responsible for mishaps (Dillenger et al., 2003). Without establishing a tone for reporting safety hazards, crew members might be reluctant to do so on their own, and may not always communicate their observations for fear of retribution (e.g.Behn et al., 1999), despite being cognizant of potential safety hazards.

Organizational Predictors. One of the more immediate work environment predictors is the organization's structure and resulting *organizational politics* (Thompson et al., 1997), which can affect perceived safety climate. It is possible that organizational politics would promote *job risk-taking.* Generally, research has found that probability of taking risks is a function of the perception of risk, appreciation of risk, likelihood of accidents/incidents, and previous outcomes (Adams, 2003; von Thaden et al. 2003). Some of the important human resource predictors affecting safety climate are *preparation and planning*, *training*, *reporting system and rewards*. *Preparation and planning* is required for safe flight operations and it has been estimated that over 100 hours of preparation are spent on each hour of flight (Sternstein & Gold, 1991). Thus, the extent to which Dispatch promotes safety as a priority consideration over financial gain might have an effect on people's perceptions of safety climate. Also, *training* efforts by an airline's management will affect perceived safety climate. An example is Crew Resource Management (CRM) training, developed in 1979, after human error was identified as the primary cause of many air transport accidents. One of the major emphases in CRM is communication of concerns, or reporting possible problems and incidents. One way airline employees are able to voice concerns is through *reporting systems*, such as NASA's Aviation Safety Reporting System (ASRS). ASRS can also be utilized for research purposes (Reynard et al., 1986) to determine safety issues and to generate safety recommendations that could eventually be implemented into FAA policies targeted towards improving safety (e.g., Burian & Barshi, 2003). Again, fear of retribution prevents people from using it (Behn et al., 1999). R*eward systems* that promote safety behavior and help to correct unsafe behaviors in an organization are needed in order to ensure a positive perception of safety climate (von Thaden et al., 2003)

Mediator and Moderators

Motivation is presented as an intermediary process variable that mediates the effects of predictors on individual (first-level) outcomes. The extent to which the stated goals are aligned with actual goals an organization is trying to reach will act as a motivator for employees to achieve the goals (Adams, 2003). Enacting stated goals for safety, thus, would likely enhance organizational safety outcomes (e.g. Griffin & Neal, 2000). According to Tuttle et al. (1975), one way to motivate employees is through performance relevant and immediate feedback, which positively affects employees' safety performance (Griffin & Neal, 2000). Thus, our framework demonstrates that the effects of various predictors, such as training, will likely affect individual outcomes, such as transfer of training, through people's motivation to achieve valued organizational outcomes, such as reduced incidents and increased well-being.

Person Moderators. Our framework postulates that certain personality and demographic variables, such as locus of control (Rochlin, 1999), propensity for risk-taking (e.g. Nicholson, 2001) and education, can moderate the relationship between safety climate predictors and safety climate outcomes.

Environment Moderators. Environment moderators identified in our framework include feedback, peer cohesion, group size, and support for safety by organizational members (i.e., management, co-worker, supervisor, and self). Previous research (e.g., Zacharatoset al., in press) has shown that feedback, peer cohesion, and support for safety are important variables that might affect safety climate in the aviation industry. Karasek and Thorell (1990) have found that job-decision latitude is associated with better work performance, positive employee attitudes, and physical and psychological well-being, whereas the opposite occurs with little decision latitude. Sadly, with increased automation, pilots sometimes see the automated flight information as a better decision-maker than themselves. Skitka et al. (1999) found that aircrews in automated conditions tended to engage in less discussion before arriving at decisions due to over-reliance on the automated systems. Peer cohesion is another potential moderator of the relationship between predictors and perceived safety climate (Simard & Marchad, 1994) and safety performance (Zacharatos et al., in press). However, excessive group cohesion may also lead to *groupthink*, which is a possible bottleneck to safety (Nicholson, 2001). Large, bureaucratic groups with dominating leaders are often reasons cited for groupthink. Thus, *group size* is a variable that might affect perceived safety climate. Another potential moderator present in the work environment is support for safety and it is important for organizations, including supervisors (Thompson et al., 1997), management, and colleagues (Fogarty, 2003; Goldman et al., 1991) to support safety initiatives.

Outcomes of Safety Climate.

Behavioral Outcomes. Behavioral outcomes often lead to organizational outcomes, such as accidents. One way to prevent accidents is to ensure *safety compliance* and minimize *risky behaviors* (Neal et al., 2000; Reason, 1997). A positive climate for safety will increase safety compliance among employees (Neal et al., 2000). Although the FAA imposes penalties for non-compliance with safety issues; if pay or other rewards are based on performance, such as on-time departures or expediting check-in, then workers might feel pressured to focus more on speed of task execution than *safety task performance* (Kaminski, 1997; Thompson et al., 1997). Because relatively few consequences are associated with inconsistent adherence to safety standards, even in the aviation industry (Thaden et al., 2003), risks are taken at the expense of passengers, crewmembers, and people in line of the flight path. Thus, poor safety climate would result in increased violations and errors (Fogarty & Neal, 2002). Violations can be prevented through *safety participation* (e.g. Goldman et al., 1991; Neal et al., 2000) and by developing *safety promoting events*, such as safety meetings that increase safety participation. Safety meetings are supposed to take place among crew members before flights, in terms of coordinating roles. Lack of crew coordination is often attributed to crew errors (Aviation Today, 2000). Unfortunately, quality of crew coordination has declined post 9/11/2001, due to new “safety” procedures (Chute, 2002).

Attitudinal Outcomes. Safety climate is expected to affect people’s attitudes, and subsequently organizational outcomes. For example, it has been noted that apathy or a bold attitude can lead to violations of safe operations and increase risk-taking (Hofmann et al., 1995). Moreover, apathy might be a result of employees becoming desensitized to safe operations over time and transferring responsibility of safety to others (Hofmann et al., 1995). That is, a poor safety climate might lead to apathetic attitudes. Also, organizational commitment (e.g. Parkes & Bochner, 2001), turnover intention, anxiety/frustration, tension, complacency, organizational/job satisfaction, safety satisfaction, and morale will be affected by perceived safety climate. In turn, these attitudes are expected to affect organizational safety outcomes. Furthermore, organizational workplace characteristics, such as communication, recognition, safety, coworkers, and feedback lead to high *morale*, which in turn, lead to job satisfaction and commitment (Fogarty, 2003). Dunbar (2001) found the extent to which employees felt management was committed to workers’ welfare and helped employees feel safe was predictive of employees’ reported satisfaction with safety in the workplace. However, with low commitment, low satisfaction, and poor safety, airline employees might report experiencing *tension*. When safety climate is perceived to be poor tension might result (Eiff & Mattson, 1998).

Cognitive Outcomes. Previous research found that exposure to informal or formal safety training and experience of incidents or accidents influences an individual’s appraisal of potential threatening situations (Goldberg et al., 1991). Furthermore, repetition of tasks leads to the ability to perform tasks with little conscious thinking regarding the steps involved (Hofmann et al., 1995), however, task performance is still subject to slips and errors (Reason, 1997). Slips or lapses are a type of cognitive error that occur due to an individuals’ dependence on memory to carry out a known task, however, the individual may depend on a wrong preexisting schema to guide execution (Hofmann et al., 1995; Reason, 1997). Therefore, in order to reduce errors, it is crucial to investigate cognitive factors (i.e., risk or situational awareness,) that result from predictors of safety climate and perceived safety climate. Safety research should also focus on sources of risk and deviations from standards (Rochlin, 1999), which are influenced by emphasis placed on representation, perception, or interpretation of risk (Krimsky & Golding, 1992) within an organization. In the aviation industry, pilots are referred to as *risk managers* to illustrate that managing risk is part of achieving goals in flight (Lofaro & Smith, 1999). Prevention of accidents can be accomplished by making sure that risk managers comprehend the gravity of risk and have the competencies for managing risks, as precursors to risk reduction (e.g. Adams, 2003). One way to ensure competencies is through reinforcement of one's *knowledge of regulations* and ensuring that off-the-job *training is transferred on-the-job*

Organization Outcomes. Organizational outcomes of safety culture and climate include attrition, accident and incident rates, reputation of safety, and employee well-being and health. The main emphasis of the aviation industry is accident prevention and a "no accident" record. Safety climate predictors, such as policies, procedures, training, and leadership (e.g., Barling et al., 2002; Burian & Barshi, 2003; Zohar, 2000), and mediators such as safety compliance and motivation (e.g., Holling, 1999) help prevent adverse outcomes (i.e., accidents, incidents, and injuries). The occurrence rate of adverse outcomes (e.g., accident rate, number of delays) can provide a measure for demonstrating the effectiveness of various safety climate predictors. In addition to physical outcomes, other social outcomes, such as a positive reputation is indicative of a positive safety culture (Schneider et al., 1994). Attrition is another organizational outcome that is influenced by climate predictors, such as the selection system of an organization. Previous research has found that mismatch of organizational and employee values, and the quality of information provided to applicants affect attrition rates (e.g. Schneider & Schneider, 1994).

Conclusion

Safety is one of the greatest demands placed on commercial airlines. However, it is not enough to have locked cockpits or to have checklists to ensure all safety procedures are followed. Airline employees must adopt a mindset for safety that ensures both procedural and common sense safety. Eiff was noted as stating, "aviation industry has been woefully negligent in addressing work-related hazards. This fact is underscored by recent exploding lost-time injury and disability claims in most aviation organizations. Increased operational tempos coupled with challenges in providing adequate staffing and equipment have generated environments rich in injury potentials" (Aviation Today, 2001, p. 3). Maintaining a safety climate is one strategy for thwarting injuries. The proposed framework exemplifies variables that might relate to perceived safety climate. Our purpose was to introduce aviation researchers to possible antecedents and consequences of safety climate. We do not recommend trying to study all these variables in one study but to study some of these variables in more simplistic models that address salient concerns.

This framework is an inclusive guide researchers and aviation practitioners can use for determining variables relevant to assessing safety climate and culture. Eventually, results of empirical research based on the framework can be molded into a tool for benchmarking safety standards across airlines. Identification of key variables related to safety culture and safety climate can enable aviation executives and safety officials to take preventative, instead of reactive, measures to enhance organizational processes that ultimately affect safety behaviors and ensure the safety and security of the flying public.

Author Note

This paper was supported by research grant No. NAG 2-1640 from NASA Ames Research Center. The authors appreciate the help of that agency.

References

Adams, C. A. (2003). *Organizational culture and safety. In The 12th International Symposium on Aviation Psychology*, (pp. 18-23). Dayton, Ohio: Ohio State University.

Aviation Today (2001, May 21). Aviation Workplace Safety Lags Other Industries. Retrieved 1/24/2003 from www.aviationtoday.com.

Aviation Today (2000, July 10). Crash shows impact on safety shortcomings on bottom line. Retrieved 1/24/2003 from www.aviationtoday.com.

Barling, J. et al. (2002). Development of a test of a model linking safety-specific transformational leadership and occupational safety. *Journal of Applied Psychology, 87*, 488-496.

Behn, L. D. et al. (1999). *Follow-up assessment of The Federal Aviation Administration's Logistics Center Safety Climate*. Report No. DOT/FAA/AM-99/19. Office of Aviation Medicine, Washington, D.C. 20591.

Bourges-Bougrine, C. P. et al. (1999). Fatigue in aviation: Point of view of French pilots. In the *Proceedings of the 10th International Symposium on Aviation Psychology*. Columbus, OH.

Burian, B. K., & Barshi, I. (2003). Emergency and abnormal situations: A review of ASRS Reports. In the *12th International Symposium on Aviation Psychology*, (pp. 176-181). Dayton, OH.

Chute, R. D. (2002, April 10). Cabin/cockpit communication: Post 9/11. Presented at the *Assault and Hijacking Workshop sponsored by the Human Factors Group of the Royal Aeronautic Society*, London Heathrow: UK.

Dillenger, T. G. et al. (2003). Relating personality with stress: Coping strategies among student pilots in a collegiate flight training program. In the *12th International Symposium on Aviation Psychology*. Dayton, OH.

Eiff, G., & Mattson, M. (1998). Moving toward an organizational safety culture. In the *Proceedings of the 1998 SAE Airframe/Engine Maintenance and Repair Conference & Exposition* (pp. 1320-1327). Long Beach, CA.

Fogarty, G. J. (2003). Errors, violations, and reporting behaviour in aviation maintenance. In the *12th International Symposium on Aviation Psychology*. Dayton, OH.

Fogarty, G., & Neal, T. (2002). Explaining safety violations and errors in the construction industry. *XXV International Congress of Applied Psychology*, Singapore.

Gadd, S. (2002). *Safety culture: A review of the literature*. Broad Lane, Sheffield: Health and Safety Laboratory, HSE.

Goldberg, A. I. et al. (1991). Threat perception and the readiness to participate in safety programs. *Journal of Organizational Behavior, 12*, 109-122.

Griffin, M. A., & Neal, A. (2000). Perceptions of safety at work: A framework for linking safety climate to safety performance, knowledge, and motivation. *Journal of Occupational Health Psychology, 5*, 347-358.

Hofmann, D. A. et al. (1995). High reliability process industries: Individual, micro, and macro organizational influences on safety performance. *Journal of Safety Research, 26*, 131-149.

Holling,H. (1999). Organizational resource management meaning, measuring, and managing. In the *10th International Symposium on Aviation Psychology*, (pp. 304-307).

Huettig, G. et al. (1999, May 3-6). Mode awareness in a modern glass cockpit attention allocation to mode information. The *International Journal of Aviation Psychology, Vol. 1. Proceedings of the Tenth International Symposium on Aviation Psychology*.

Kanki, B. G. (1996). *Enhancing the flight safety culture through training*. Moffett Field: NASA Ames Research Center.

Karasek, R., & Theorell, T. (1990) *Healthy work: Stress, productivity and the reconstruction of working life*. New York: Basic Books.

Lazarus, R. S., & Folkman, S. (1984). *Stress appraisal and coping*. New York: Springer

Lofaro, R. J., & Smith, K. M. (1999). Operational decision-making (ODM) and risk management (RM): Rising risk, the critical mission factors. In the *Proceedings of the Tenth International Symposium on Aviation Psychology*.

Mearns, K. J., & Flin, R. (2001). Assessing the state of organizational safety--Culture or climate? In H. Ellis & N. Macrae (Eds.), *Validation in psychology* (pp. 5-20). New Brunswick, NJ: Transaction Publishers.

Mearns, K. et al. (2001). Benchmarking safety climate in hazardous environments: A longitudinal, interorganizational approach. *Risk Analysis, 21*, 771-786.

Nicholson, N. (2001). An evolutionary perspective on change and stability in personality, culture, and organization. In M. Erez, U. Kleinbeck, & H. Thierry (Eds.), *Work motivation in the context of a globalizing economy* (pp. 381-392). New Jersey: Erlbaum.

Orlady, H.W. & Orlady, L.M. (1999). Aviation human factors: A "core technology" for flight operations. In the *Proceedings of the 10th International Symposium on Aviation Psychology*. Dayton, OH

Parkes, L. P., & Bochner, S. (2001). Person-Organization fit across cultures: An empirical investigation of individualism and collectivism. *Applied Psychology: An international Review, 50*, 81-108.

Reynard, W. D. et al. (1986). *The development of the NASA Aviation Safety Reporting System* (NASA Technical Report no. 1114). Moffett Field: NASA Ames Research Center.

Rochlin, G. I. (1999). Safe operation as a social construct. *Ergonomics, 42*, 1549-1560.

Schneider, B. et al. (1994). Creating a climate and culture of success. *Organizational Dynamics, 23*, 17-29.

Schneider, B., & Schneider, J. L. (1994). Biodata: An organizational focus. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 423-450). Palo Alto, CA: CPP Books.

Simard, M., & Marchad, A. (1994). The behavior of first-line supervisors in accident prevention and effectiveness in occupational safety. *Safety Science, 17*, 169-185.

Skitka, L. J. et al. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, *51*, 991-1006.

Sternstein, E., & Gold, T. (1991). *From take-off to landing.* New York: Pocket Books.

Thompson, R. C. et al. (1997). *Where the safety rubber meets the shop floor: A confirmatory model of management influence on workplace safety*. FAA Civil Aeromedical Institution, Oklahoma City, OK.

Tuttle, T. G. et al. (1975). Organizational Psychology. In B. L. Margolis & W. H. Kroes (Eds.), *The human side of accident prevention: Psychological concepts and principles which bear on industrial safety* (pp. 7-46). Springfield, IL: Charles C. Thomas.

von Thaden, T. L. et al. (2003). Safety culture in a regional airline: Results from a commercial aviation safety survey. In the *12th International Symposium on Aviation Psychology*. Dayton, OH.

Zacharatos, A. et al. (in press). *High performance safety systems and Occupational Safety*. School of Business, Queen's University.

Zohar, D. (2000). A group level model of safety climate: Testing the effects of group climate on microaccidents in manufacturing jobs. *Journal of Applied Psychology, 85*, 587-596.

Zohar, D. (2002). Modifying supervisory practices to improve subunit safety: A leadership-based intervention model. *Journal of Applied Psychology, 87*, 156-163.

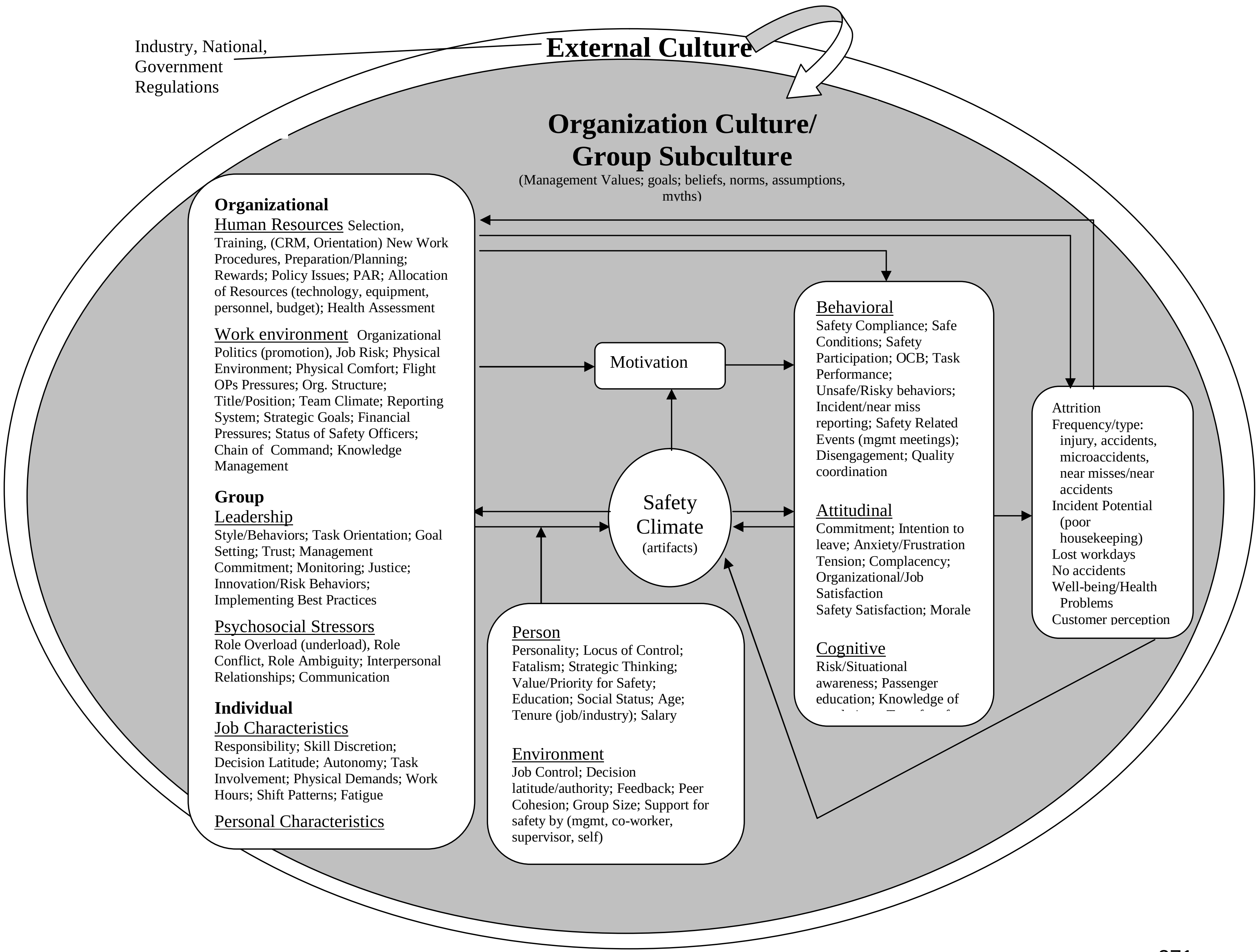


Figure 1: *Framework for Studying Safety Climate in the Aviation Industry*

EFFECTS OF WORKLOAD AND VISIBILITY ON MISSION REHEARSAL

Tessa Gorton
Tokyo, Japan
t_gorton@hotmail.com

Gavan Lintern
General Dynamics
Dayton, Ohio 45431-1289
glintern@earthlink.net

Mission rehearsal poses new opportunities and new challenges for flight simulation. The general issue, how to promote transfer to the criterion task, is the same for mission rehearsal as it is for training. On the other hand, the goal of mission rehearsal is to promote sensitivity to or awareness of contextual details that are crucial to success of a specific mission while the goal of training is to develop generic skills. It is not clear, at this stage, what implications these different goals have for the design of simulators. For the navigation mission examined here we hypothesized that high workload and restricted visibility would distract attention from important navigation information and thereby slow development of navigation knowledge. Both experimental manipulations had the hypothesized effect under some experimental conditions but not under others. The differential effectiveness of the manipulation under different conditions offers some insight into the nature of the navigation-relevant information that can be enhanced by mission rehearsal.

Introduction

Within aviation and other technological work environments where operators control complex systems, simulators have found use primarily as devices for teaching or maintaining general skills. Mission rehearsal (familiarization of an experienced operator with a specific task scenario) offers a different opportunity for simulators to enhance operational performance. The use of mission rehearsal during the conflict in Bosnia (Defense Mapping Agency, 1997) demonstrates a perceived need in the operational community. In addition, there is a class of relatively common aviation incidents that can be characterized as misinterpretations by experienced pilots due to unfamiliarity with specific contextual details of a mission (Bone, 1997).

Workload

Underlying the interest in mission rehearsal is the belief that unfamiliarity with certain specific details of a scenario can disrupt smooth progress through that mission. From this perspective, a pilot who is rehearsing a mission should be given the opportunity to attend to those specific details. High workload is one feature of a rehearsal that might prevent that. Lintern and Wickens (1991) have reviewed data which suggest that high workload on one task can impede learning of another concurrent task.

The most direct evidence of the impact of workload on learning is from studies by Nissen and Bullemer (1987) and Lindberg and Garling (1982). Nissen and Bullemer (1987) demonstrated that the learning of response pattern could be slowed by a concurrent secondary task. Lindberg and Garling (1982) similarly showed that a concurrent secondary task could slow learning of distance and direction judgments in a simplified navigation task. The body of research in this area is vulnerable, however, to the criticism that the tasks considered were, at best, simplified abstractions of real-world tasks.

As a means of exploring the workload issue in this experiment, order of roll control was manipulated during rehearsal. Some subjects rehearsed on a system with first-order (velocity) roll dynamics during familiarization sessions and others rehearsed on a system with second-order (acceleration) roll dynamics. We assumed that by changing the roll dynamics in this manner we would change the difficulty of basic control. Subjects with second-order roll dynamics should have to pay more attention to flight control and would thereby have their attention diverted from navigation. We hypothesized that this would degrade the effectiveness of rehearsal.

From one perspective, this sort of manipulation falls into the category of a difficulty manipulation. There is considerable uncertainty expressed in the literature regarding the effects of transfer from easy to difficult and difficult to easy tasks (Holding, 1961; Lintern, Roscoe, & Sivier, 1990). In this study we chose to examine both directions of transfer by having all subjects fly systems with first- and second-order roll dynamics in separate (and counterbalanced) transfer trials. By this strategy we were able to assess the effects of transfer from easy to difficult and from difficult to easy conditions relative to the appropriate control conditions of easy-to-easy and difficult-to-difficult transfer respectively.

Visibility

There is uncertainty about the type of scenario-specific information that is crucial to a successful mission. In a navigation task of the type used in this experiment, it may be landmark knowledge, which is knowledge of specific details of a route to be followed (Golledge, 1991; Hirtle & Hudson, 1991). On the other hand, it may be survey knowledge, which is knowledge of layout and of relationships between features (Hirtle & Hudson, 1991). It is also possible that different types of navigational challenges will impose burdens on different types of knowledge. For example, a straight route between two waypoints may be supported effectively by recognition of individual features while a winding course between two waypoints may require a better sense of global spatial relationships.

This issue was examined by manipulation of visibility in rehearsal. The criterion mission was to navigate the route under restricted visibility. Some subjects rehearsed the route with unrestricted visibility and others with the mission level of restricted visibility. Transfer to the less difficult condition of unrestricted visibility was of some interest but was not tested in this experiment because the transfer trials already incorporated a rather complex set of issues.

We propose that rehearsal with restricted visibility will be advantageous if specific navigational features on or near the course are important but that rehearsal with unrestricted visibility will be advantageous when information some distance off course is needed for learning the spatial layout of the course. Individual legs varied in characteristics we hypothesized to be important. Some legs were rich and others poor in landmark and route knowledge and, while most legs were straight, one wound through a series of hills. We hypothesized that the winding leg and also legs poor in landmark and route knowledge would benefit from rehearsal with unrestricted visibility during rehearsal because that condition would offer subjects more opportunity to become attuned to off-course information.

Method

Subjects

The experimental design called for 48 subjects. Four whose runs resulted in missing data due to system crashes and another who did not return for a scheduled session were replaced.
Forty-eight pilots (34 male and 14 female) completed the experiment. All were working towards a private pilot license in the pilot training program at the University of Illinois and, as a result, had some prior navigational training. Individual levels of flight experience ranged between 30 and 120 hours with a median of 46 hours. All subjects had 20/20 vision or better (corrected or uncorrected) and were aged between 18 and 31 years.

Apparatus and Stimuli

Visual imagery was generated at an update rate of 50Hz with an Evans and Sutherland (E&S) SPX 500T image generator and projected by two Electrohome ECP 3000 color projectors onto screens each measuring 304.8 cm by 228.6 cm. The right-hand screen was placed directly in front of the viewpoint and the left-hand screen to the left at an angle of 115 degrees for a viewing angle of 112 x 38 degrees (27 degrees right to 85 degrees left of the centerline) at a viewing distance of 300 cm. An offset to the left was used because all waypoint turns but one were to the left. Consequently, most of the critical navigational information was located either straight ahead or to the left of the current heading.

Flight instruments were generated by an IRIS Silicon Graphics Computer and displayed in a head-down location on a separate monitor. Heading was displayed at the top of the screen in both analogue and digital forms. The analogue display gave subjects a better sense of direction of the turn while the digital display supported more precise judgments. Altitude above ground (AGL) was presented on a vertical analogue scale along the right side of the screen. A moving arrow on a stationary scale showed altitudes ranging from 0 to 200 feet. Above 200 feet the pointer went to the top of the scale and the altitude was represented digitally at the top of the screen in blue. The target altitude of 150 feet was represented in white and the rest of the scale was drawn in black. The attitude indicator showed a fixed aircraft symbol on a rotating artificial horizon with a pitch ladder. It also provided a measure of bank angle. No other flight parameters were presented.

Subjects sat in a chair directly in front of the simulation with the joystick mounted on the right arm. This joystick was a two-axis Flightstick, which gave control of pitch and bank angle. The bank angle was limited to 30 degrees and power was preset to maintain airspeed at approximately 85 knots. Yaw was preset at zero degrees for all trials and there was a six- to eight-knot crosswind directly from the left on all legs.

The Navigation Task

The simulated navigation area was approximately 13.5 by 13.5 nautical miles (Figure 1). The topography of the area included both flat and hilly terrain with rivers, roads and buildings. For this experiment, a low-fidelity version of the area was used for a rehearsal phase and a high-fidelity version was used for a transfer phase. Objects were distributed along the course to ensure that there were always one or more features in view to guide navigation. In the high-fidelity version of the area, these features had distinctive characteristics but care was taken to ensure that they were not placed directly on course (where they might have been used as indicators of direction to the next waypoint). The low-fidelity version of the area contained the same objects as in the high-detail world but differed in the appearance of those objects. Hills appeared to be more block-like than those of the high detail world and objects such as buildings and bridges were represented as gray blocks. In the development of this low-fidelity version, the intent was to use a level of detail that would be available with a less capable image generation system.

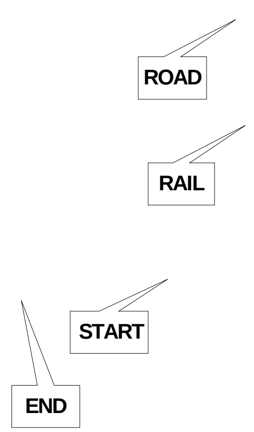


Figure 1. *The navigational area used for the rehearsal and transfer flights (The depicted course was shown only to the guided rehearsal groups and then only during their rehearsal flights)*

The course had seven legs of 38 nautical miles (nm) total length (individual legs ranged from 3.7 to 5.0 nm). The range in altitude of this course was 750 feet. As a secondary task, subjects were to maintain an altitude of 150 feet above ground level, which required vertical speeds of approximately ± 1500 fpm in the climbing and descending portions of the route. The course could be completed in approximately 27 minutes. A 6- to 8-knot variable crosswind from the left and light turbulence in pitch and roll were present to make the task more challenging. This ensured that the subjects were prevented from simply memorizing leg headings.

An automatic procedure was programmed to reset subjects to the start point for the next leg (with heading aligned with the course of that leg) if elapsed time for the current leg was 30% greater than a criterion time. That criterion time had been established from the time taken by an experimenter to fly that leg with the course clearly indicated by a line superimposed on the scene. A message appeared on the monitor towards the end of each leg to advise subjects either that they had reached the end of the leg or that they were being reset.

A different navigation area was used to familiarize subjects with the simulation. It had a five-leg course that required approximately 15 minutes to complete. A path was marked on the ground to guide subjects along the designated course. Turns, climbs and descents were similar in magnitude to those required to navigate the course laid out in the other area.

Experimental Factors

Visibility. Visibility was either unrestricted (nominally fifteen miles) or restricted by haze (two miles). Unrestricted visibility permitted a view beyond the end of each leg (in the absence of physical obstructions) from the start point of that leg and also a view of features well to the side of the course. In contrast, restricted visibility only permitted a view of the area in close proximity to the current position. At the start of the first leg, for example, unrestricted visibility permitted a clear view of an upcoming mountain but with restricted visibility only a white haze was visible until the road and railway line came into view.

Control Stability. A first-order control system was used to implement a high-stability condition and a second-order control system was used to implement a low-stability condition.

Procedure

Three experimental phases (system familiarization, rehearsal and transfer) were run in sequence over two experimental sessions of two hours each. In the first session there were two familiarization trials (with the familiarization navigation area) and two low-fidelity navigation trials. The second session was scheduled either one or two days later. It started with the final low-fidelity familiarization trial. Two high fidelity transfer trials followed.

Familiarization. Subjects were familiarized with the control dynamics of the flight simulation. The primary task was to fly directly over a path marked on the terrain and the secondary task was to maintain an altitude of 150 feet AGL. Each subject completed two familiarization flights, the first with a first-order control system and the second with a second-order control system. Subjects were advised of the change in control order and of how that would change the task. Light turbulence in pitch and roll and a variable 6- to 8-knot crosswind directly from the left were included on both flights. After the completion of each flight, vertical and horizontal root mean squared (RMS) errors were displayed on the monitor. The meaning of these errors was explained to subjects.

Rehearsal. The rehearsal phase used the low-fidelity navigation area. Subjects were randomly assigned to one of four groups encompassing two levels of visibility and two levels of control stability. The task was to navigate along the predetermined course. Subjects were given a map of the area on which the route was clearly marked and were advised that they should fly this route by relating map symbols to landmarks shown in the simulation. As a secondary task, subjects were required to maintain an altitude of 150 feet AGL as in the familiarization session.

Transfer. The transfer phase followed the third rehearsal trial. The path was the same as flown in rehearsal. Subjects were given five minutes to study the map. They then flew the course twice without the map. Visibility was set at two miles (the restricted level used in familiarization) for both trials. Half of the subjects flew with first-order control first and half flew with second-order control first.

Dependent Measures

Lateral deviations from course and vertical deviations from the target altitude were measured for individual legs from the start of each leg up to a point 2000 feet from the endpoint of that leg. These errors were converted to RMS error scores.

Analyses

Each leg for both the rehearsal and transfer sessions was analyzed separately. Partial correlations between lateral and vertical performance measures were examined for the rehearsal and the transfer sessions to assess the feasibility of conducting univariate tests on the between-subjects effects. These correlations were at least moderately high in general (0.45+) thereby indicating that univariate tests would not be appropriate (Tabachnick & Fidell, 1989).

MANACOVAs were used to test the statistical significance of effects for the dependent measures of lateral and vertical error and for trials (three for rehearsal and two for transfer). RMS error performance on the familiarization trials was used as a covariate. The analyses conducted on the rehearsal data included multivariate tests of the trials effects and also of interactions with trials. Significant multivariate effects from tests on the combined transfer trials were followed by separate MANACOVAs on each of the two transfer trials to assess effects of the training factors on transfer performance at each level of stability.

Note for the discussion of results that, while it is tempting to consider horizontal and vertical errors separately, the logic of multivariate analysis (as supported by high correlations between the two measures) does not permit that. Error scores must be considered a unitary Horizontal-Vertical dimension.

Results

Figures 2-4 show mean horizontal RMS errors (transformed to their natural logarithms) for the second, fifth and seventh legs of the three rehearsal and two transfer flights.

Rehearsal Trials

There were significant effects for stability on all legs: lamba (6,38) for leg 1 = 8.99, $p < 0.001$; for leg 2 = 17.19, $p < 0.001$; for leg 3 = 14.22, $p < 0.001$; for leg 4 = 7.12, $p < 0.001$; for leg 5 = 9.90, $p < 0.001$; for leg 6 = 7.74, $p < 0.001$; for leg 7 = 7.20, $p < 0.001$. There were also significant effects for visibility on the first, sixth, and seventh legs, lamba (6,38) for leg 1 = 3.80, $p < 0.005$; for leg 2 = 2.19 $p < 0.065$; for leg 3 = 1.15, $p < 0.35$; for leg 4 = 2.03, $p < 0.09$; for leg 5 = 1.94, $p < 0.10$; for leg 6 = 2.78, $p < 0.02$; for leg 7 = 4.52, $p < 0.002$. Performance was better for the stable versus the unstable system and for unrestricted versus restricted visibility (Figures 2-4).

Transfer Trials

Significant effects of the Stability manipulation were found in rehearsal on the stable transfer trial in legs 2, 5, and 7, lamba (2,42) for leg 2 = 6.95, < 0.002; for leg 5 = 6.30, p< 0.004; for leg 7 = 4.31, p< 0.02 (Figures 2-4). Rehearsal on Low Stability led to poorer control in transfer to the stable system. There were no main effects of the Stability manipulation in rehearsal on the unstable transfer trial.

Significant effects of the Visibility manipulation were found in rehearsal on the stable transfer trial in legs 2 and 7, lamba (2,42) for leg 2 = 3.95, < 0.027; for leg 7 = 5.82, p< 0.006 (Figures 2 & 4). Only in Leg 7, where Low Visibility led to poorer control in transfer to the stable system, were the trends sufficiently clear to interpret. There were no main effects of the Visibility manipulation in rehearsal on the unstable transfer trial.

A significant interaction of Stability by Visibility was found in rehearsal on the unstable transfer trial in leg 5, lamba (2,42) for leg 5 = 2.77, < 0.07.

Discussion

Stability

The stability manipulation was introduced in rehearsal to test the hypothesis that high workload would divert attention from the navigational task to the control task. Under these circumstances, subjects should pay less attention in rehearsal to features on and near the course that would assist their navigation in the subsequent transfer trials. The rehearsal data indicate that this manipulation did affect the difficulty of the task. On all legs, rehearsal performance was better with the stable system.

The differential effects of rehearsal stability on the transfer trials were confined to the stable transfer trial of legs 2, 5, and 7. Use of stable control in rehearsal led to better performance on the stable transfer trials. Although this result is consistent with our workload hypothesis, it is also consistent with the popular high fidelity hypothesis.

Visibility

Effects of the visibility manipulation were evident only for legs 2 and 7. The visibility effects for Leg 2 cannot be interpreted with confidence, but the effects for Leg 7 show that subjects who rehearsed with unrestricted visibility performed better on the stable transfer trial. This is of particular interest because the transfer trials were run under the restricted visibility condition. Any high-fidelity conceptualization of transfer would predict that rehearsal with restricted visibility would be advantageous. In contrast, this result is consistent with our hypothesis that high visibility rehearsal would reveal information that could then be used effectively in a low visibility mission.

Conclusion

In this project, we have added to the somewhat meager data that show that high workload in training can disrupt learning. Furthermore, in contrast to those other data, we have shown this effect with a more complex and more realistic task.

In contrast to general training, which has the goal of developing generic skills, mission rehearsal seeks to promote sensitivity to or awareness of contextual details that are crucial to success of a specific mission. It is common to assume that high fidelity in rehearsal will ensure good mission performance. Here we challenge that assumption and show that the high-fidelity assumption does not account consistently for the data.

The contrasting hypothesis, following Lintern (1991), is that conditions that permit an operator to pay attention to critical mission details are more likely to develop the specific skills needed to accomplish a successful mission. In consideration of mission rehearsal effectiveness, fidelity is a spurious and bankrupt construct. There is now considerable evidence that no form of fidelity or similarity theory (whether physical or psychological) can account for important transfer effects (e.g., see Lintern, 1991). Notions of fidelity and similarity serve only to distract from exploration of the real issue, that being the specific type of manipulations that can make mission rehearsal effective.

Acknowledgments

The Basic Research Office of the Army Research Institute under Contract MDA 903-93-K-0006 supported preparation of this article. Dr. Michael Drillings was the technical monitor. We are grateful for the generosity of the Evans and Sutherland Corporation, Salt Lake City, who donated the SPX image generation system to the University of Illinois.

References

Bone, R. S. (1997). Active participation versus map study: Geographic knowledge acquired through three forms of aviation mission rehearsal. Thesis submitted in partial fulfillment of requirements for Master of Arts, University of Illinois, Champaign, Illinois.

Defense Mapping Agency. (1997). DMA fact sheet: New system lets pilots "fly" over Bosnia before takeoff. [Web page]

Golledge, R. G. (1991). Cognition of physical and built environments. In T. Garling and G. W. Evans (Eds.), Environment, cognition, and action: An integrated approach. Oxford: Oxford University Press.

Hirtle, S. C., & Hudson, J. (1991). Acquisition of spatial knowledge of routes. Journal of Environmental Psychology, 11, 335-345.

Holding, D. H. (1961). Transfer between difficulty and easy tasks. British Journal of Psychology, 53, 397-407.

Lindberg, E., & Garling, T. (1982). Acquisition of locational information about reference points during locomotion: The role of central information processing. Scandinavian Journal of Psychology, 23, 207-218.

Lintern, G. (1991). An informational perspective on skill transfer in human-machine systems. *Human Factors, 33*, 251-266.

Lintern, G., Roscoe, S. N., & Sivier, J. E. (1990). Display principles, control dynamics, and environmental factors in pilot training and transfer. Human Factors, 32, 299-317.

Lintern, G., & Wickens, C. D. (1991). Issues for acquisition and transfer of timesharing and dual-task skills. In D. Damos (Ed.), Multiple task performance. (pp. 123-138). London: Taylor & Francis.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. Cognitive Psychology, 19, 1-32.

Tabachnick, B. G., & Fidell, L. S. (1989). Using multivariate statistics. New York: Harper & Row.

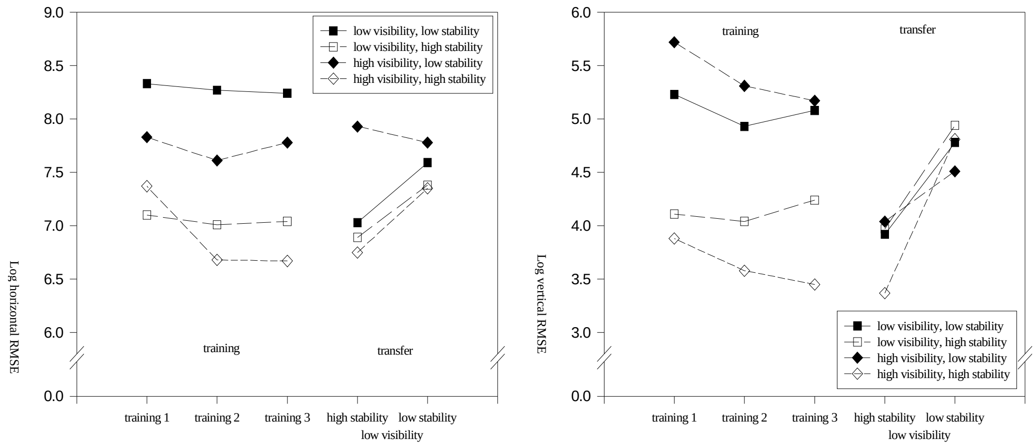


Figure 2: *Mean RMS errors (transformed to their natural logarithms) for the 2nd leg of the three rehearsal and two transfer flights*

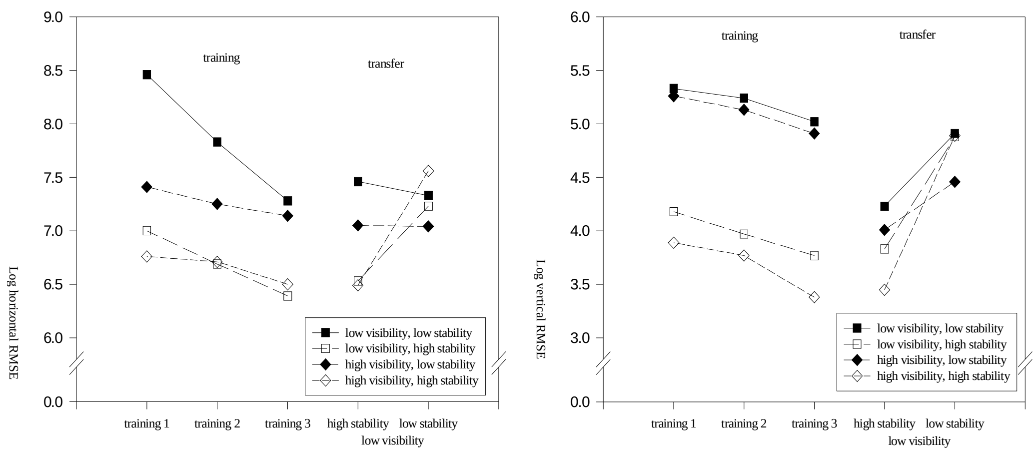


Figure 3. *Mean RMS errors (transformed to their natural logarithms) for the 5th leg of the three rehearsal and two transfer flights*

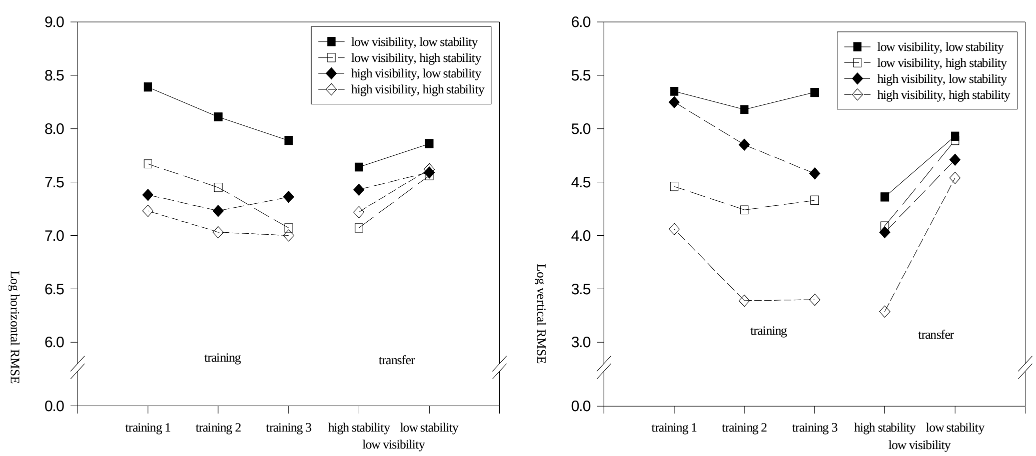


Figure 4. *Mean RMS errors (transformed to their natural logarithms) for the 7th leg of the three rehearsal and two transfer flights*

DEVELOPMENT AND INTEGRATION OF A HUMAN-CENTERED VOLUMETRIC COCKPIT SITUATION DISPLAY FOR DISTRIBUTED AIR-GROUND OPERATIONS

Stacie Granada and **Arik Quang Dao**
San Jose State University
Moffett Field, CA

Dominic Wong
QSS Group Inc.
Moffett Field, CA

Walter W. Johnson and **Vernol Battiste**
NASA Ames Research Center
Moffett Field, CA

In the Distributed Air Ground Traffic Management (DAG-TM) simulation environments pilots flew desktop simulators, which included a Cockpit Situation Display (CSD). Within the current paper we will briefly review the tasks pilots were responsible for in the simulations and subsequently evaluate the tools made available on the CSD to assist the pilots in executing their tasks. Some of the tasks pilots were responsible for in the simulations included the following: to create and evaluate user-preferred routes, meet flight scheduling requirements at the meter fix, self-space behind designated aircraft, and maintain separation with other aircraft. Some of the tools offered within the CSD to facilitate these tasks included a Route Analysis Tool (RAT), a Waypoint table with capabilities to input scheduling requirements, a Spacing tool, and Conflict Detection and Alerting logic. A detailed examination of these features and others will be discussed

In 1995, the RTCA Task Force 3, Free Flight Implementation (1995) cited a need for a Cockpit Display of Traffic Information (CDTI) that could increase situation awareness on the flight deck in order to develop and progress the notion of free flight. While numerous definitions of free flight have surfaced, a seminal view expressed in the final report was that *any* move toward removing restrictions on behalf of the user is a step toward free flight. In order to support free flight, the task force suggested that a CDTI needed to provide information that would allow the flight deck to maintain separation with other aircraft, perform rerouting operations en route, and engage in limited delegation to maintain spacing en route or in the terminal area.

The NASA Ames Flight Deck Display Research Laboratory has devoted many years of research and development to a Cockpit Situation Display (CSD; a high fidelity aviation navigational display) due to the increase in endorsements toward advanced flight deck displays by the FAA and NASA's Advanced Air Transportation Technologies (AATT) program (Johnson, Battiste, & Holland, 1999). Work on the Ames 3D CSD described in this paper was conducted for two reasons. First, there is an accepted need for displays that provide vertical and horizontal situation information. In an assessment of both these display formats, Wickens, Olmos, Chudy, & Davenport (1997) found that information about relative altitude was not naturally available when viewing traffic on a horizontal situation display, and lateral position information was not available on vertical situation displays. In today's flight deck where display space is at a premium, having a single display to view both vertical and horizontal situation information seems to be the practical solution.

Second, an advanced flight deck display was needed to facilitate the examination of potential free flight concepts. In order for Air Traffic Controllers (ATC) to manage higher traffic flows they will not only need advanced tools on their end, but they also will need pilots to share in some of the air traffic management roles and responsibilities. By providing a tool such as a CSD on the flight deck pilots can maintain better situation awareness, which inadvertently gives ATC greater flexibility with what options they can use to manage their own workload, such as with having pilots manage to a required time of arrival at the meter fix. Overall, an advanced flight deck system is needed to facilitate the examination of free flight concepts.

Multiple simulations were conducted at NASA Ames Research Center to examine Distributed Air Ground Traffic Management (DAG-TM) solutions for free flight. These proposed solutions aimed to examine three concept elements - CE-5: En Route Free Maneuvering; CE-6: En Route Trajectory Negotiation; CE-11; Self-Spacing for Merging and In-trail Separation for Terminal Arrival - which were designed to evaluate various roles and responsibilities by which free flight could be achieved (Battiste et al., 2005; Johnson et al., 2005). The goal of the current paper is to review a Decision Support Tool (DST; specifically, the Ames 3D CSD) utilized by pilots in these simulations. The main tools developed within the Ames 3D CSD are described in detail. We address some of the benefits of the Ames 3D CSD and the tools within the CSD, both of which provided the platform with which we could test the concepts proposed under DAG-TM.

The CSD and DAG-TM

The goal of DAG-TM was to propose a prototype of an air/ground system with a human-centered approach. That is, the research team reevaluated the roles and responsibilities of the stakeholders to enhance user flexibility and user efficiency with, for example, user-preferred routing to increase airspace capacity without impeding upon safety or airspace accessibility.

Within the concept elements tested, pilots flew desktop simulators and were responsible for the following tasks: 1) maintaining separation 2) meeting their assigned RTA (Required Time of Arrival) 3) modifying Ownships flight path for traffic and RTA compliance 4) sending and acknowledging trajectory changes, and 5) self-spacing behind a designated lead aircraft. Several DSTs were provided to aid pilots in accomplishing these new tasks and to meet their responsibilities in the simulations. The remaining portion of this paper addresses the Ames 3D CSD, which provided pilots with the ability to achieve the tasks outlined above.

CSD Display Overview

The primary DST for the flight deck was the 3D CSD (see Figure 1), which dynamically depicted traffic, flight plans, conflicts, and more. With this airside interface, pilots had the ability to view traffic information in a planar view, profile view, and to dynamically position the display in some combination between these two choices with a 3D perspective. An earlier paper describes some of the features of a previous 2D version of the CSD in detail (Johnson, Battiste, & Holland, 1999).

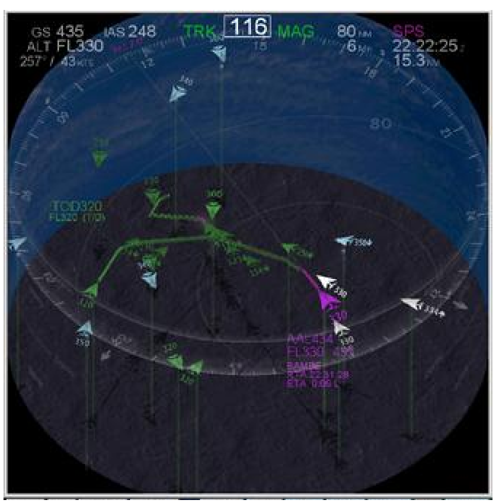


Figure 1. *3D CSD; to view pictures in color or to download a demo version see the web address in footnote 1.*

The goal of the Ames 3D CSD was to integrate several tools into a single interface and expand the user's situation awareness by providing a 3D depiction of the airspace. 3D displays have some advantages compared to 2D displays. First, 2D planar displays do not visually render altitude information in as optimal a manner as was desired. It is possible to use coplanar displays exhibiting both top down and profile views of traffic, but they use excessive display space and cannot depict some conflict geometries. It was intended that pilots have the ability to view any and all geometric traffic situations, thus the display needed to have the ability to depict any and all geometric traffic situations. Second, future work with the current CSD will include the integration of traffic, weather, and terrain within the same display. 3D renderings may provide more realistic depictions of weather and terrain, as well as provide greater global situation awareness within the flight deck.

CSD - Display Basics. Due to limited space, only the key components of the DSTs within the Ames 3D CSD are described in this paper[1]. In general, the Ames 3D CSD presents the standard navigational elements at the top-most portion of the display and Ownship is depicted as magenta.

The Ames 3D CSD offers two modes in which a user can view traffic. At start-up, the primary display projection (the standard view) is set in *Expanded mode, planar view*. In this mode a compass rose, depicted at the top-most portion of the CSD, displays 100 degrees of heading value with Ownship depicted at the center of the display. In the *Expanded mode*, the user can only view the CSD in a 2D top-down (i.e., planar) view. When the display is switched to *Full mode*, the user can manipulate the CSD to examine the display as a 3D depiction on a 2D surface (i.e., perspective display). In this mode, the compass rose is displayed as a full 360 degrees around Ownship (except in temporal view, which is described below).

The benefit of providing the Full mode in the Ames 3D CSD is that two projection views are offered: *Orthographic* and *Perspective*. From the orthographic view, the sizes and perspective of elements on the screen are discrete and constant which makes it easier for the viewer to make judgments regarding the distance and direction of the aircraft on the display. In contrast, the perspective view affords making relative judgments regarding the

[1] For more information, download the Ames 3D CSD User Guide at http://human-factors.arc.nasa.gov/ihh/cdti/download.html

relative distance between aircraft; as aircraft that are farther away diminish in size.

In addition to orthographic and perspective views, the full mode provides a *Central* and *Temporal* view which are dependent upon the relative position of Ownship. The *Central* view positions Ownship in the center of the rings on the display affording detection of aircraft by a distance parameter around the Ownship. The display range provides the ability for pilots to zoom in for a detailed look at the airspace around Ownship (10nm) or zoom out as far as 640nm to facilitate in viewing or planning far-term trajectories. Table 1 lists the display range tool and indicates how usable and useful pilots in the simulations found this tool.

The *Temporal* view positions Ownship closer to the bottom edge of the display, which maximizes the view in front of Ownship. The depiction of traffic is relative to a time parameter, where for example, any aircraft displayed can reach Ownship within 10 minutes.

The Ames 3D CSD provides four memory settings whereby users can set and quickly flip through multiple views. For example, a user may choose a top-down planar view of the traffic, a vertical rear-view, a vertical side-view (or profile view), and a 3D view. With visual momentum, the display will move into any of the preference settings by simply clicking the corresponding buttons. Additionally the display can be manipulated into any 3D view by simply right clicking and dragging the display toward the desired angle. Although it is not likely that a mouse will be used on a flight deck to manipulate such a display, this input device works for simulations and other control devices can eventually adopt similar strategies for acquiring display motion. Further research is needed to explore this issue as implementation of CSDs come closer to reality.

Usefulness of the display. Research shows that there are advantages and disadvantages to 3D displays and 2D coplanar displays, and the benefits of each are task dependent (Wickens, Olmos, Chudy, & Davenport, 1997). The 3D CSD outlined here has the benefit of being manipulated to display a top down view of the traffic situation, a profile view, or *dynamically* moved to display some rare conflict geometries that are not discernable from simple 2D or coplanar displays. Having the option to view traffic from several viewpoints allowed pilots to look ahead at any conflict situation and determine where the paths of two aircraft would cross while searching for an efficient route through the meter fix. In the DAG-TM simulation of autonomous flight operations, pilots flying in an Advanced Concepts Flight Simulator (ACFS) and those flying single-pilot stations (both using Ames 3D CSDs) were able to meet their meter fix crossing restrictions while maintaining separation with other aircraft (Kopardekar et al., 2004). Pilots reported using the Ames CSD in 3D 36% of the time and in 2D 64% of the time. Table 1 provides pilots' ratings of features within the Ames CSD in terms of usability and usefulness (each item is addressed in the text).

Table 1. *Pilot Ratings of CSD Tools*

Tool	Usability		Usefulness	
	M	SD	M	SD
Display settings				
Display range	4.9	0.32	5.0	0.00
3D View	4.2	1.14	4.1	1.29
Route information				
3D flight plans	4.3	1.34	4.7	1.00
Path predictors	5.0	0.00	4.7	0.67
RAT features				
RAT path	4.5	0.53	4.6	0.70
RAT: drag/drop	4.5	0.85	4.6	0.84
Alerting system				
Alert warning	4.5	0.71	4.5	0.71
Alert symbology	4.4	0.52	4.0	1.05

N = 10; Scale: 1 = not very usable/useful, 5 = very usable/useful.

Pilots found the display range, 3D views, and 3D flight plans usable and useful. Overall, the 3D display settings provided pilots with enough situation awareness to maintain separation and make strategic flight modifications, and the pilots seemed to like these features.

CSD - Aircraft Intent. With the Ames 3D CSD, properly equipped aircraft (i.e., those with Automatic Dependent Surveillance Broadcast, ADS-B) have the ability to transmit and receive intent information, or flight plan information, from aircraft within ADS-B broadcast range. The Ames 3D CSD allows users to view intent information in two ways. First a user may choose to display the entire flight plan of one or more aircraft on a case-by-case basis. Flight plan intent information is depicted as a linear path relative to the direction of the aircraft heading and includes information regarding level flight and descent or ascent segments of flight. To view the flight path of an aircraft, a user must simply click on the aircraft symbol within the CSD and the flight path is rendered on the display. Again, for specific details regarding intent depiction see the Ames 3D CSD User Guide (http://human-factors.arc.nasa.gov/ ihh/cdti/ download.html).

The second option to view intent is for users to display "fast-time predictor" information simultaneously for *all* or selected aircraft within the selected altitude surveillance band and broadcast range. Fast-time predictors are shown along the depicted flight path of all or selected aircraft. The users can select the amount of time they would like the intent information "forecasted". That is, traffic may be displayed with path intent information ranging from 0 to 20 minutes ahead. The fast-time predictor is shown as a pulse traveling the length of the predictor line(s) and reflects the speed the aircraft are flying. Due to the nature of flight, intent information may change at any moment. Intent information is depicted for the "current" status of the aircraft and updates as the flight status for each aircraft changes.

The pulse predictor also traverses any flight plan (option 1 from above) that is selected by the user as long as the pulse predictor is set anywhere from 2 – 20 minutes. The major distinction here is that flight plan information yields *all* of the aircraft's registered intent information, whereas the predictor lines only show up to 20 minutes of intent information. Again, the flight plan of individual aircraft may be turned on or off by clicking on the desired aircraft symbol, then the predictor can be turned on for the selected aircraft, or on for all aircraft.

Usefulness of aircraft intent. There are benefits associated with the predictor tool and having access to visualizations of entire flight plans. For example, Xu and Rantanen (2003) demonstrated that perceptual cues regarding motion prediction afford less error in collision estimations. This supports the notion that the predictor tool can offer robust situation awareness in detecting conflicts as it provides perceptual information regarding future locations of the target aircraft. That is, combined with a 3D display the pulse predictor provides 4D flight information, which fosters low workload for examining the threat potential of existing traffic.

Additionally, as users have access to rendering of full flight plan information they are less likely to fall subject to the perceptual illusions. Research has demonstrated that there are particular geometric collision angles that may elicit a bias in position prediction and create a false sense of safety when predictor lines are short (Comerford & Uhlarik, 2001; Holland, 1998). Therefore, the full flight plan implementation allows the user to scrutinize possible conflicts more closely for safer operations. As Table 1 indicates, pilots found the flight plans and pulse predictor tool highly usable and useful.

CSD - Route Assessment Tool (RAT). The RAT provides the user with the ability to create and visualize in-flight route modifications, submit proposed route modifications to ATC, receive route modifications from ATC, and execute any of these modifications depending on flight status. The planning and implementation of these flight plan modification possibilities are subject to provisional alerting and are made available for 1) strategic conflict resolution, 2) RTA requirements, 3) weather avoidance, 4) direct route efficiency, 5) dynamic Special Use Airspace (SUA) avoidance, and eventually 6) terrain avoidance.

To perform any of the functions outlined above, the user must first turn the RAT on by clicking the RAT button on the CSD toolbar. This provides the user with access to a waypoint table, a flight path that can be manipulated (see Figure 2), and options for execution or datalink. The RAT tool allows the user to enter new waypoints or to use the waypoint table to scroll through existing waypoints.

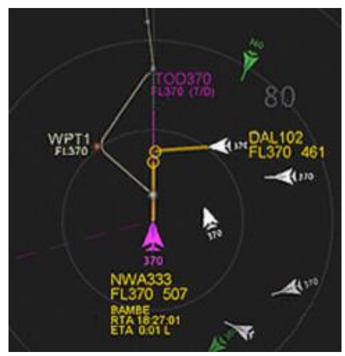


Figure 2. *Route modification with RAT*

Once a waypoint is identified, the user can enter an RTA for the waypoint, enter a new altitude for that waypoint, or move the waypoint to a new lateral position. All of the RAT functions can be visualized and evaluated before execution, which helps reduce the need to make numerous changes to the flight plan since pilots can verify whether the modification reflects the desired action.

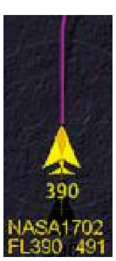
Usefulness of the RAT. In the recent DAG-TM simulations, the RAT allowed pilots to solve conflicts strategically as opposed to tactically. That is, pilots were able to modify their flight plans to avoid a loss of separation (Kopardekar et al., 2004), whereas with existing TCAS systems, collision threats can only be avoided tactically when the threat is imminent.

The RAT provided an easy method for visualizing, manipulating, and changing flight information, which required little mental effort or calculations on behalf of the flight crew. For example, if a pilot wanted an altitude change, the only work required by the crew was to insert a start point on the flight path and to enter the newly desired flight level. The aircraft's most economical climb or descent is determined by the system and is based on how large the change in altitude is. Since the RAT provides immediate visual feedback and provisional alerting information, flight crews know whether the proposed flight change will have an adverse impact on safety before executing the plan, and can continue to visually search for path variations until a safe trajectory is found. In the recent DAG-TM simulations, pilots were able to use the RAT to strategically solve conflicts when the traffic levels exceeded the capacity of today's airspace (Kopardekar et al., 2004). Table 1 also indicates that pilots found the design of the RAT path and the drag and drop features of the RAT to be usable and useful.

CSD - Alerting. The Ames 3D CSD alerts are depicted for strategic conflict detection as opposed to tactical conflict detection. This type of alerting is designed to encourage less drastic changes to the flight plan in order to resolve the conflict to help reduce time, cost, and to increase safety. The alerting logic detects conflicts (or losses of separation) based on an algorithm of temporal proximities, which takes into account the aircraft intent information or aircraft state information (current heading, altitude, speed). For more detailed information regarding logic behind the Conflict Detection and Resolution within the CSD see Canton, Refai, Johnson, and Battiste (2005).

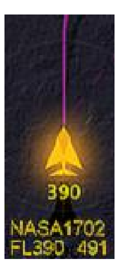
The CSD depicts 3 levels of alert (See Figure 3a, b, c). Alert level 1 is the lowest level of alert and is depicted on the CSD when Ownship and the conflicting aircraft become yellow or amber. At Alert Level 2, an amber glow is added to the existing Alert Level 1 symbology. Finally at Alert Level 3, yellow predictor lines with intersecting Loss of Separation (LOS) rings are added to the alert symbology. These depictions provide information regarding how imminent any particular alert may be.

Figure 3a



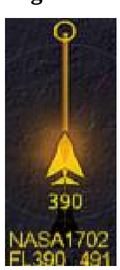
Alert Level 1

Figure 3b



Alert Level 2

Figure 3c



Alert Level 3

Usefulness of alerting. The alerting techniques utilized by the CSD were useful in providing enough information to keep pilots aware of possible safety concerns (i.e., possible losses of separation) without committing to numerous false alarms (Kopardekar et al., 2004). Xu (2003) recommends that effective alerting systems provide continuous measures of conflict detection as opposed to dichotomous measures. That is, rather than provide pilots with an all or nothing view of whether a conflict is likely, it is beneficial to utilize and consider the dynamics of flight. Again, the alerting system outlined here took into account winds, future flight plan information (such as a descent profile), and the alert level based on proximities. Additionally pilots had access to the time-to-contact information (with early notification), which overall contributed to the pilots' ability to view possible conflicts at farther time increments, allowing strategic resolutions rather than tactical. As with the other CSD tools, Table 1 indicates that pilots found the alert warning and alert symbology usable and useful.

CSD - Spacing. The CSD allows users to self-space behind designated lead aircraft (e.g., maintain 90 seconds behind aircraft XYZ). With the spacing tool users can input the assigned spacing value while algorithms work to adjust Ownship's speed in order to maintain the required interval (Abbott, 2002).

The CSD renders a spacing box that represents the target location of Ownship based on the spacing interval that was set. This provides the user with updated visual information regarding the current spacing status. For example, if Ownship is targeted at the correct interval behind its lead aircraft, the spacing box will appear green and Ownship will visually appear *in* the box. Similar visual feedback is provided if Ownship is too early or late for its spacing assignment. A temporal indication of the spacing status also appears in Ownships data tag when the spacing is set.

Usefulness of spacing. In the DAG-TM simulations, pilots were able to effectively use the spacing tool and they found workload to be low when spacing clearances were issued early (Battiste et al., 2005). With the spacing tool available on the Ames 3D CSD, it is possible to test several concepts aimed at improving airspace bottlenecks as aircraft transition from en route through the meter fix into the terminal area.

Ames 3D CSD Effectiveness

Overall, the Ames 3D CSD utilized in the DAG-TM simulations provided pilots with the type of situation awareness necessary to effectively maintain separation, meet their assigned RTA, modify Ownships flight path, send/acknowledge trajectory changes, and self-space behind designated lead aircraft.

The tools within the Ames 3D CSD offered a flexible and comprehensive backbone of information that pilots could use in testing the autonomous operations, self-spacing spacing operations and more. In this simulation we had the opportunity to evaluate how pilots interacted with the aforementioned tools and how they in turn facilitated flight in this futuristic free flight concept (Kopardekar et al., 2004). The data indicated that flight crews flying the ACFS and pilots flying the single (desktop) station CSDs were able to meet their assigned RTA's as well as the speed and altitude restrictions at the meter fix, whether they were under ATC control or operating autonomously. It is also worth noting that an increase in traffic did not alter the CSD pilots' performance in meeting these requirements. Finally, all CSD pilots were able to maintain separation with both the managed and autonomous aircraft, even with the increase in traffic. This demonstrates the potential for free flight concepts with the use of CSDs, such as the one described here.

The Ames 3D CSD presented here has incorporated visually dynamic traffic information, such as aircraft intent, route planning, and conflict alerting. Future work will address the integration of traffic, weather, and terrain on a 3D display. Preliminary work on incorporating weather into the CSD is currently undergoing investigation, and recommendations have been made regarding possible design issues to consider for this type of integration (Comerford, 2004).

References

Abbott, T. (2002). *Speed Control Law for Precision Terminal Area In-Trail Self Spacing*. NASA Technical Memorandum 2002-211742. Hampton, VA: National Aeronautics and Space Administration.

Battiste, V., Palmer, E., Johnson, W., Smith, N., Prevot, T., Mercer, J., Granada, S., Johnson, N., Dao, Q., Lee P. (2005). Human Centered Decision Support Tools for Arrival Merging and Spacing. *Submission to the 2005 International Symposium on Aviation Psychology*, Oklahoma City, OK

Comerford, D. (2004). *Recommendations for a cockpit display that integrates weather information with traffic information*. (NASA Technical Memorandum 2004-212830).

Comerford, D., & Uhlarik, J. (2001). The arc-size illusion as applied to planar displays of aircraft traffic information. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*. Santa Monica, CA: HFES.

Holland, S. (1998). Effect of predictor length on conflict detection in a cockpit display. *Thesis presented to faculty at San Jose State University.*

Johnson, W. W., Battiste, V., Granada, S., Johnson, N., Dao, Q., Wong, D., and Tang, A. (2005). A Simulation Evaluation of a Human-Centered Approach to Flight Deck Procedures and Automation for En Route Free Maneuvering. *Submission to the 2005 International Symposium on Aviation Psychology*, Oklahoma City, OK

Johnson, W. W., Battiste, V., and Holland, S. (1999). A cockpit display designed to enable limited flight deck separation responsibility. *Proceedings of the 1999 World Aviation Conference*, Anaheim, CA.

Kopardekar, P., Battiste, V., Johnson, W., Smith, N., Palmer, E., Barhydt, R., Palmer, M., Wing, D., Lee, P., Prevot, T., Mercer, J., Mafera, P., Hitt, J., Doble, N., and D'Arcy, J. F. (2004*). Distributed Air-Ground Traffic Management Concept Element 5: En Route Free Maneuvering Joint Ames/Langley Experimental Evaluation of Integrated Air/Ground En Route Operations.* (Milestone Report 8.503.7 Deliverable Submitted to the AATT Project Office).

Radio Technical Commission for Aeronautics (RTCA). (1995b). *Free flight implementation* (RTCA Task-Force 3 Report). Washington, D.C.: RTCA Inc.

Canton, R., Refai, M., Johnson, W., and Battiste, V. (2005). Development and integration of human-centered conflict display and resolution tools for airborne autonomous operations. *Submission to the 2005 International Symposium on Aviation Psychology,* Oklahoma City, OK

Wickens, C. D., Olmos, O., Chudy, A., & Davenport, C. (1997). *Aviation display support for situation awareness.* (Logicon Technical Report ARL-97-10/LOGICON-97-2)

Xu, X. (2003). Conflict detection with cockpit display of traffic information: What is it, what has been found, and what needs to be done*? Proceeding of the 47th Annual Meeting of the Human Factors and Ergonomics Society.* Santa Monica, CA. HFES.

Xu, X., and Rantanen, E. M. (2003). Timing of discrete actions: Evidence for perceptual dominance in human performance. *Proceeding of the 47th Annual Meeting of the Human Factors and Ergonomics Society.* Santa Monica, CA.

THE COSTS AND BENEFITS OF HEAD-UP DISPLAYS (HUDS) IN MOTOR VEHICLES

Lisa Hagen, **Matthew Brown**, **Chris M. Herdman**, and **Dan Bleichman**
Centre for Applied Cognitive Research
Carleton University
Ottawa, Canada

Research in the aviation domain has shown that Head-Up Displays (HUDs) can facilitate performance in specific tasks such as controlling aircraft flight path and altitude (Fadden, Ververs, & Wickens, 2001; McCann & Foyle, 1995; Martin-Emerson & Wickens, 1997; Wickens & Long, 1995). However, there are a number of simulator-based studies suggesting that pilots may focus, or *cognitively tunnel* their attention on HUD symbology, resulting in performance decrements in tasks that require continuous monitoring of information from the outside scene (Foyle, Stanford, & McCann, 1991; McCann, Foyle, & Johnston, 1993), and in extreme cases, severe impairment or even failure to detect potentially critical discrete events in the external scene (Brickner, 1989; Fischer, Haines, & Price, 1980; Wickens & Long, 1995). In the present research, we extended our examination of aircraft HUDs to the domain of motor vehicles. Participants drove a high fidelity, fully configured driving simulator through a realistic scenario containing both urban and rural (highway) roads. Speed limit (and other) signs were posted. Two conditions were compared. In the no-HUD condition, a standard in-vehicle instrument panel was used. In the HUD condition, the instrument panel was augmented with a HUD showing digital speed on the windshield. The results showed a benefit of the HUD insofar as participants were better at maintaining their speed in the HUD than in the no-HUD condition. However, this benefit was accompanied by a cost in that participants showed significantly greater deviations in maintaining lane position when the vehicle's speed was available on the HUD than when it was not. This finding suggests that HUD symbology distracts motor vehicle operators to the extent that they are less able to process information from the navigation environment.

Introduction

HUD technology, traditionally used in aircraft, has been implemented by various automobile manufacturers to project vehicle status information onto the windshield (e.g., speed, warning lights). Although there is thorough research on the efficacy of HUDs in aircraft, relatively little work has been done on the impact of HUDs in motor vehicles.

In the present research, the impact of a digital HUD speedometer on driving performance was assessed using a high-fidelity driving simulator. To quickly preview the results, the present study shows that although this particular HUD improved a driver's ability to monitor speed, it impaired their ability to maintain lane position. This trade-off is explained in terms of cognitive tunneling.

Theoretical Benefits and Costs of HUDs

The benefit of HUDs, whether they are implemented in aircraft or in cars, is that they allow the user to monitor vehicle status without *physically* interfering with their ability to view the navigation environment. In theory, HUDs should provide the driver with more time to attend to events in the navigation environment. However, findings from studies testing the effects of HUDs in aircraft suggest otherwise (e.g., Herdman & LeFevre, 2003; McCann & Foyle, 1995). These studies showed that pilots have more difficulty detecting objects/events in the navigation environment when HUD information is available, relative to when it is not. One explanation for this counter-intuitive finding is that pilots are susceptible to a cognitive tunneling effect when a HUD is available. That is, the HUD symbology captures (and holds) the pilot's attention, subsequently preventing them from attending to other events in the navigation environment.

Cognitive Tunneling and HUDs in Automobiles

It seems plausible that the inherent costs and benefits of HUD technology observed in aircraft operation would map directly onto the task of driving an automobile. However, the navigation environment faced by pilots is sparsely populated relative to that faced by a typical driver. As such, drivers are required to navigate in environments that require more precise control of their vehicle's position both within lane markings and relative to other cars sharing the lane. It may therefore be the case that the cognitive tunneling effects observed in flight simulation studies (see Herdman & LeFevre, 2003; McCann & Foyle, 1995) are relatively minor both in terms of magnitude and in terms of consequence. The ever-increasing number of HUDs being installed in automobiles magnifies the importance of assessing the (a) extent to which HUDs render drivers susceptible to cognitive tunneling and (b) the subsequent impact on driving performance.

In order to determine whether cognitive tunneling occurs in automobile HUDs, a simulation experiment was conducted in which drivers' performance in terms of their ability to monitor speed and lane position was assessed. The critical (within-subjects) manipulation had two conditions: (1) participants used the manufacturer-equipped analogue speedometer to ascertain speed (no-HUD condition) and (2) the analogue speedometer was augmented with a HUD of a digital speedometer (HUD condition). The participants' driving performance in these two conditions was compared to determine whether HUD information yields costs and/or benefits.

Methods

Participants. Twenty-two Carleton University students participated and either received course credit or $20 remuneration. All participants were assumed to have normal or corrected-to-normal vision. Further, all participants held a valid Province of Ontario driver's license and had at least two years of driving experience.

Design. One critical factor with two levels was manipulated (HUD condition: HUD vs. no-HUD). This factor was counterbalanced across participants such that half received the HUD condition first and the no-HUD condition second. This order was reversed for the other half of the participants.

Apparatus. The experiment was conducted on a high-fidelity, fully configured DriveSafetyTM 500c driving simulator consisting of a (partial) cabin of a Saturn passenger car mounted in front of five flat-screen projectors subtending approximately 22° of vertical visual angle and 150° of horizontal visual angle. The HUD information (i.e., a digital display of the vehicle's current speed) was located 5° of visual angle below the horizon and 10° of visual angle to the left of the center of the driver's field of view. The HUD was light green in color and subtended 4° of visual angle vertically and 2° of visual angle horizontally. Computer-generated engine noise, which changed accordingly with engine speed, and external noise (e.g., passing traffic) were presented on speakers mounted in the cabin or on the cabin platform. The driving scenario was scripted using TCL scripting language that was executed on a PC-based Linux platform and simulated a two-lane highway passing through small towns, mountain passes, and rural farming areas. The scenario was updated at a rate of 30 to 60 Hz and the data were collected at a rate of 5 Hz.

Procedure. Participants familiarized themselves with the controls and operation of the driving simulator during a ten-minute practice session. The HUD was displayed during practice to minimize potential novelty effects associated with its presence during the experimental session. The experimental session consisted of two identical 25-minute trials, except that participants used the HUD to monitor their speed (HUD condition) on one trial, whereas they used the analogue speedometer on the other (no-HUD condition). Participants were instructed to (a) obey all posted speed limits and general rules of the road and (b) keep the vehicle centered in their lane. Participants were debriefed and received appropriate compensation following completion of the second experimental trial.

Results

Two participants were removed from the analyses: one was unable to complete the experiment due to illness and the other misunderstood task instructions. The data from the remaining 20 participants were trimmed such that data at both the beginning and at the end of the experiment (accelerating to the posted speed limit and decelerating to a full stop) were eliminated. Outlier data were eliminated based on the criteria that the participant's lane position deviated 1.8 m (or more) from the center of their lane.

Speed Monitoring Data

Participants' ability to monitor speed was measured by comparing actual speed to the posted speed limits. This measure was calculated by taking the absolute value of the difference between their actual speed and the speed limit. Speed monitoring was significantly better in the HUD condition than in the no-HUD condition $t(19) = 9.0$, $p < .001$. On average, speed in the no-HUD condition deviated from the speed limit by 3.98 MPH, whereas it only deviated by 2.48 MPH in the HUD condition.

Lane Position Data

The ability to monitor lane position is a critical aspect of safe driving, given that the consequences of failing to do so (e.g., crossing into oncoming traffic) are so dire. Indeed, it could be argued that lane position monitoring is more important that speed monitoring in terms of road safety. For this reason, participants' lane position data were logged and subsequently analyzed. The center-most position of the lane was assigned a value of zero and any deviation left of center was recorded as a negative value, whereas deviations right of center were assigned a positive value. Although knowing about possible systematic tendencies to drift in one direction relative to the other could be of some interest, it is beyond the scope of the present research. As such, lane position monitoring performance was calculated by taking the absolute value of their current lane position (which represents the distance from the center of the lane given that the center of the lane was assigned a lane position value of zero). The interesting result here is that lane position monitoring was significantly *worse* in the HUD condition than in the no-HUD condition $t(19) = 4.3$, $p < .001$. On average, participants in the HUD condition drifted .33 m from the center of their lane, whereas participants in the no-HUD condition only drifted .29 m from the center of their lane. This difference of .04 m could represent the difference between a “close call” and a head-on collision.

Discussion and Conclusion

The results from this driving simulation experiment show that when a manufacturer-equipped analogue speedometer is augmented with a digital speed HUD, drivers are better at monitoring their speed, but *worse* at maintaining their lane position, relative to when no HUD is available. These results are consistent with the claim that digital speed HUDs (typical of HUDs used in automobiles) render participants susceptible to cognitive tunneling effects, whereby attention is captured and held by the HUD symbology such that it is difficult (or impossible) to concurrently attend to information in the navigation environment (e.g., lane position).

Although monitoring vehicle speed is important, the consequences of failing to do so pale in comparison to the potentially disastrous outcomes of neglecting one’s lane position or not being able to detect objects and/or events in the navigation environment (e.g., a child running into the roadway). As such, the present research suggests that the limited benefits of a digital speed HUD are outweighed by the potential costs associated with not adequately processing information in the navigation environment. It is therefore essential to refine and empirically assess how and what information (if any) should be presented on a HUD so as to maximize driver awareness of vehicle status while minimizing potential cognitive tunneling effects.

References

Brickner, M.S. (1989). Apparent limitations of head up displays and thermal imaging systems. In R.S. Jensen (Ed.), *Proceedings of the Fifth International Symposium on Aviation Psychology* (pp. 703-707). Columbus, OH: The Ohio State University.

Fadden, S., Ververs, P., & Wickens, C. (2001). Pathway HUDs: Are they viable? *Human-Factors*, 43, 173-193.

Fischer, E., Haines, R. F., & Price, T. A. (1980). Cognitive issues in head-up displays. *NASA Technical Paper 1711*, Ames Research Center, Moffett Field, CA.

Foyle, D. C., Sanford, B., & McCann, R. S. (1991). Attentional issues in superimposed flight symbology. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 577-582). Columbus, OH: The Ohio State University.

Herdman, C. M. & LeFevre, J. (2003). Situation awareness, performance and workload in simulated flight: Head-up display (HUD) versus head-down display (HDD). *Government of Canada Report No. 007SV-W7714-010547*.

Martin-Emerson, R. & Wickens, C. (1997). Superimposition, symbology, visual attention, and the head-up display. *Human-Factors*, 39, 581-601.

McCann, R. S. & Foyle, D. C. (1995). Scene-linked symbology to improve situation awareness. *AGARD Conference Proceedings CP 575* (16-1 - 16-11). Brussels, Belgium.

McCann, R. S., Foyle, D. C., & Johnston, J. C. (1993). Attentional limitations with head-up displays. In R. S. Jensen (Ed.), *Proceedings of the 37th Annual Meeting of the Human Factors Society* (1345-1349). Columbus, OH: The Ohio State University.

Wickens, C. & Long, J. (1995). Object versus space-based models of visual attention: Implications for the design of head-up displays. *Journal of Experimental Psychology: Applied*, 1, 179-183.

Acknowledgements

We are grateful to Transport Canada for providing access to a driving simulator for this research.

Correspondence should be directed to:
Dr. Chris M. Herdman, Centre for Applied Cognitive Research, Carleton University, Ottawa, ON. Canada, K1S 5B6.

Direct email correspondence to:
chris_herdman@carleton.ca

THE DEVELOPMENT OF AN INTELLIGENT HELICOPTER PERFORMANCE INSTRUMENT SYSTEM INCORPORATING HUMAN FACTORS AND SKILLED BASED EXPERTISE

Michael Hardy
Flight Test Instructor
National Test Pilot School
Mojave California

Steven Thatcher
Aviation Education, Research and Operations Laboratory/
Knowledge-based Intelligent Engineering Systems Centre
School of Electrical and Information Engineering
University of South Australia,
Adelaide, Australia

Studies by Iseler and De Maio (2001) suggest that the helicopter accident rate in the USA is 10 times that of airline accidents, with 20% attributed to pilot error. US Army/NASA studies (Hart 1998) have developed a number of strategies with the goal of reducing the helicopter accident rate including real-time prediction, measurement and pilot cueing. Hardy and Thatcher (2004) described early development work in designing an intelligent helicopter performance instrument system that conducts performance predictions in real time that is being pursued with the aim of providing pilot cueing indicating sufficient performance is or is not available for takeoff, landing and maneuvering flight phases. This paper further describes the research work completed to date including the developmental instrument system that has been assembled from commercial-off the shelf hardware and software that will allow rapid prototyping from results of in-flight performance and human factors testing. The helicopter performance instrument system could significantly reduce pilot workload and enhance safety. It has the potential to assist in reducing the accident rate associated with collisions with the ground during takeoff and landing attributed to the pilot being unaware of insufficient performance of the helicopter.

Introduction

Analyses of US civil rotorcraft accident records over the last two decades by Iseler and De Maio (2001) suggest that the helicopter accident rate to be some 10 times that of corresponding airline accidents, with 20% of the accidents attributed to pilot error, with the primary factor associated with the skill level of the pilot. The highest accident rates in helicopters has been shown to occur with lower cost helicopters and is likely related to the lower skill level of the pilot. Lower skill pilots are more often involved in accidents that their own error is a primary or major factor. As a result of their studies, the US Army/NASA Rotorcraft Division reported through Hart (1998) that they have developed a number of strategies to attain a goal of reducing “the rotorcraft accident rate attributable to human factors and drive train malfunctions by a factor of five by the year 2007”. Amongst these strategies are through the use of human-centered cockpit technology that includes ”real-time aiding to eliminate inadvertent envelope exceedance with real time prediction, measurement, cueing and limiting for critical parameters and components”.

Although several modern helicopters already provide some means of performance prediction through embedded Flight Management Systems (FMS), generally these are limited in their capability to predict hover capability and provide a power assurance check, as described by Eurocopter EC130B4 Flight Manual (n.d.) and the Boeing AH-64D Operator’s Manual (n.d.). These are also aircraft specific and not adaptable to other helicopter types.

This paper presents the development work completed to date in developing an intelligent helicopter performance instrument system that conducts performance predictions in real time and provides pilot cueing for takeoff, landing and maneuvering flight phases. The system uses consumer Pocket PC devices running applications developed using National Instruments LabVIEW software that allow rapid prototyping to address results from performance and human factors testing.

Engine Power Available Versus Power Required

In order to conduct a safe takeoff, landing or maneuvering flight, sufficient engine power must be available to meet or exceed the helicopter power required to conduct the maneuver. Engine power available is dependent upon pressure altitude, ambient temperature and engine condition. Helicopter power required for a task is also dependent upon pressure altitude and ambient temperature, as well as variables such as airspeed, gross weight, center of gravity, vertical acceleration forces, height above the ground, attitudes, size and frequency of control inputs, and aircraft configuration. The difference between power available and power required is termed power margin. Provided the power margin is equal or greater than zero, the pilot has some assurance the sufficient power is available to conduct the maneuver.

The engine power available may be power turbine temperature, gas generator speed or torque limited. Exceeding any of these limitations may result in engine or transmission damage, or the engine may have insufficient power to maintain rotor speed once past these limits, with subsequent loss of thrust and anti-torque effectiveness. The limit which is reached first will be dependent upon the ambient conditions. For example in cold, low altitude conditions, the gas turbine engine typically will be able to produce more power to the transmission than the transmission rating and so will reach a torque limit first. In high temperature, high altitude conditions, the engine may reach a power turbine temperature limit prior to the torque limit; or perhaps the gas generator speed limit prior to the turbine temperature or torque limit.

The power required to conduct a specific task, whether it be a takeoff, landing or maneuvering flight, is dependent upon how much thrust is required from the main rotor and tail rotor, plus power required to run accessories (hydraulics and electrics), as well as mechanical losses. Predicting the power required for a maneuver is not a trivial task. This can only be easily achieved for steady state conditions such as hover or level flight. The Federal Aviation Administration Regulations, 14 CFR Part 27 (2003) and Part 29 (2003) require manufacturers to provide data in the Rotorcraft Flight Manual (RFM) which the pilot can use to predict whether sufficient power is available at varying conditions to conduct certain tasks such as a hover; or data to predict takeoff distances for transport category helicopters.

However the high workload presented to a helicopter pilot, particularly for helicopters without stability augmentation or single-piloted, often leads to the pilot not referring to RFM performance data as a flight progresses. Not doing so means the pilot either resorts to guesswork; previously calculated predictions with large safety margins; or really has no idea if sufficient power is available to safely complete a maneuver. Guesswork as an alternative to performance planning has been a contributing factor in many helicopter accidents, often classified as insufficient planning or poor pilot judgment. Typically the helicopter either has insufficient power to complete a take off, colliding with the ground or obstacles; or, insufficient power to arrest the rate of descent during landing or maneuvering. This is particularly an issue when the takeoff or landing site is at a considerable altitude or temperature difference to where the flight originated, or when significant gross weight changes have occurred (extra fuel, cargo or passengers). Applying excessively large safety margins are also undesirable as they can result in reduced productivity of the helicopter with the pilot electing to carry a lighter payload than actually achievable.

Existing Helicopter Performance Predicting Human-Centered Cockpit Technology

Several modern helicopters already provide some means of performance prediction, engine power assurance checks and display through embedded Flight Management Systems that relieve the pilot from the workload associated with referring to the RFM during flight. These include Eurocopter's Vehicle and Engine Management Display system used in the EC Series of helicopters such as the EC130 (Eurocopter EC130B4 Flight Manual, n.d.) and the performance planning system integrated in the Boeing AH-64D Apache Longbow (Boeing AH-64D Operator's Manual, n.d.). However these are limited in their ability to predicting maximum hover weight and base their computations on a minimum engine power specification. Should the engine be producing more or less power than the minimum specification, the performance predictions in these systems will be inaccurate. These systems are also aircraft-type specific and not adaptable to other helicopters.

Developmental Instrument System

In order to address the limitations of existing performance prediction instrumentation systems and provide real-time performance prediction and cueing for hover, takeoff, landing and maneuvering, a developmental instrument system for integration into a test helicopter and in-flight evaluation has been developed. It consists of commercial off-the-shelf hardware and software in the form of two IPAQ h5150 Pocket PC devices running application code written using National Instruments LabVIEW and LabVIEW PDA Module software. One of the Pocket PC's holds a National Instruments DAQCard 6062E

PCMCIA data acquisition card in an expansion pack. This Pocket PC is designed to acquire atmospheric, helicopter engine and aircraft performance data in real time and transmit the parameters via a Bluetooth link to the second Pocket PC on the pilot's kneeboard for computation and display. The Bluetooth connectivity leaves the pilot and PDA unencumbered by wires and simplifies aircraft integration. While the graphics capable of being displayed for pilot cueing by LabVIEW PDA Module software are somewhat limited, this approach will allow rapid prototyping in response to performance and human factors tests that will be undertaken to evaluate workload, display formats, moding and automation. The system has been bench tested (Figure 1) and is currently being integrated into an ex-US Army Bell OH-58C helicopter, operated by the National Test Pilot School, Mojave, California (Figure 2).

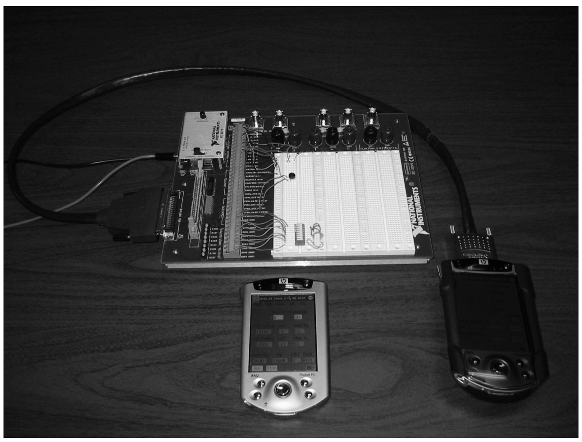


Figure 1. *IPAQ/LabVIEW Based Intelligent Helicopter Performance Instrument System under Bench Test*

Model and Display Development

The first step taken has involved developing detailed performance models for the test helicopter based on the performance data from the OH-58C Operator's Manual (US Army, 1989), US Army flight test reports (Benson, Buckhanin, Mittag and Jenks, 1975, and Spring, Buckhanin, Burch and Niemann, 1979), and takeoff performance predicting methods developed by Burke, Schmitz and Vause (1977).

Figure 2. *OH-58C Test Helicopter*

Code has been developed from these models to predict and display on the Pilot's Pocket PC engine power assurance data, maximum engine power available, power required to hover, hover ceiling, takeoff distances to clear obstacles and available power margins. In particular, power margins between engine power available and predicted power required to hover in ground effect and out of ground effect are displayed and updated in real time based on pressure altitude, outside air temperature and helicopter gross weight inputs. Further to this, knowledge-based techniques are currently being investigated, with the aim of developing a prediction agent based on radial base function neural networks outlined by Thatcher, Jain and Fyfe (2004) and Haykin (1999); or a multilayered perceptron neural (Alonso-Betanzos, Fontenla-Romero, Guijarro-Berdinas, Hernandez-Pereira, Canda, Jimenez, Legido, Muniz, Paz-Andrade and Paz-Andrade, 2002) that would yield a satisfactory prediction of successful optimization of take off, landing and maneuvering events. These techniques will consider actual power being produced by the helicopter engine, rather than just the minimum engine specification. One must also consider that experts (pilots in this case) bring a vast amount of experience or knowledge to the process and the use of expert knowledge as input to the knowledge-based agent/s is also being considered, as suggested by Van Aartrijk, Tagliola and Adriaans (2002).

Conclusions

This paper summarizes the development work completed to date for an intelligent helicopter performance instrument system that conducts performance predictions in real time and provides pilot cueing for takeoff, landing and maneuvering flight phases. The system uses commercial off-the-shelf Pocket PC devices running applications developed using National Instruments LabVIEW software that will allow rapid prototyping to results from human factors testing. The system has the potential to significantly reduce pilot workload and enhance safety. In particular it has the potential to assist in reducing the accident rate associated with collisions with the ground during takeoff and landing attributed to the pilot being unaware of insufficient performance of the helicopter.

References

Amparo Alonso-Betanzos, Oscar Fontenla-Romero, Bertha Guijarro-Berdinas, Hernandez-Pereira, Juan Canda, Eulogio Jimenez, Jose Legido, Susana Muniz, Cristina Paz-Andrade and Maria Paz-Andrade, A Neural Network Approach for Forestal Fire Risk Estimation.,*15th European Conference on Artificial Intelligence*, Lyon, France, pp643-647, 2002.

Burke, LTC J.A., Schmitz Dr. F., Vause, R. Optimizing Takeoffs Of A Heavily Loaded Helicopter, US Army Air Mobility R & D Laboratory, *US Army Aviation Digest*, May 1977.

Benson, T., Buckhanin, R., Mittag, C., and Jenks, J., United States Army Aviation Engineering Flight Activity Edwards Airforce Base, Evaluation of an OH-58A helicopter with an Allison 250-C20B Engine, USAAEFA Project No. 74-48, dated April 1975

Boeing AH-64D Apache Longbow Flight Manual, Chapter 10, Section 7.1-7.3.5.

Eurocopter Flight Manual EC130B4, original issue.

Federal Aviation Administration, DOT, 14 Code of Federal Regulations Part 27, *Airworthiness Standards: Normal Category Rotorcraft*; 2003.

Federal Aviation Administration, DOT, 14 Code of Federal Regulations Part 29. *Airworthiness Standards: Transport Category Rotorcraft;* 2003.

Hardy, M., and Thatcher S J., (2004). Development of a Helicopter Performance Instrumentation System. In Negnevitsky, M. (ed.), *The Proceedings of the 2nd International Conference on Artificial Intelligence in Science and Technology (AISAT'2004.)* (2004), Tasmania: University of Tasmania

Hart, S, Army/NASA Rotorcraft Division, Analysis of Civil Helicopter Accidents, seminar presented at *HeliExpo '98*, Feb 15, 1998.

Haykin, S., *Neural Networks: a Comprehensive Foundation.* Prentice Hall International, 1999.

Iseler and De Maio, Analysis of U.S. Civil Rotorcraft Accidents from 1990 to 1996 and Implications for a Safety Program, *American Helicopter Society*, May 2001.

Operator's Manual Army Model OH-58C Helicopter, Headquarters, Department of the Army, TM 55-1520-228-10, 17 Jan 89.

Spring, S., Buckhanin, R., Burch, T., and Niemann, J., United States Army Aviation Engineering Flight Activity Edwards Airforce Base, Airworthiness and Flight Characteristics Evaluation OH-58C Interim Scout Helicopter, USAAEFA Project No. 76-11-2, dated April 1979.

Thatcher, S. J., Jain, L. C. and Fyfe, C., (in press). An Intelligent Aircraft Landing Support System. In *Lecture Notes in Computer Science: The Proceedings of the 8th International Conference on Knowledge-based Intelligent Information & Engineering*., Berlin: Springer-Verlag, 2004

Van Aartrijk, Tagliola and Adriaans., "AI on the Ocean: the RoboSail Project, *15th European Conference on Artificial Intelligence*, Lyon, France, pp653-657, 2002.

EFFECTS OF A FINAL APPROACH RUNWAY OCCUPANCY SIGNAL (FAROS) ON PILOTS' FLIGHT PATH TRACKING, TRAFFIC DETECTION, AND AIR TRAFFIC CONTROL COMMUNICATIONS

John Helleberg
The MITRE Corporation
McLean Virginia

Eighteen pilot participants with varying experience levels flew 36 approaches in a medium fidelity cockpit simulator. Eighteen baseline trials were flown with a standard Precision Approach Path Indicator (PAPI) and 18 trials were flown with the proposed Flashing PAPI (FPAPI). The results showed a significant increase in lateral tracking error with the FPAPI as compared to the PAPI trials, but no increase in vertical tracking errors. There was also a trend toward an increase in the number of radio communications with the FPAPI. Pilots were able to determine runway occupancy status and land or go-around as required in both the baseline and FPAPI trials.

Introduction

On February 1, 1991 at Los Angeles International Airport (LAX), USAir flight 1493 (USA1493), a Boeing 737, was landing on runway 24L when it collided with Skywest flight 5569 (SKW5569), a Fairchild Metroliner, which was positioned at an intersection awaiting clearance for takeoff on runway 24L. As a result of the collision, both airplanes were destroyed. All 10 passengers and 2 crewmembers aboard SKW5569 were killed, as were 20 passengers and 2 crewmembers aboard USA1493 NTSB (1991). As this and other recent accidents have shown runway incursions pose a significant safety risk.

At present, there is no automated capability in the National Airspace System (NAS) to directly warn airborne flight crews of runway occupancy status at either controlled or uncontrolled airports. The Final Approach Runway Occupancy Signal (FAROS) concept was designed to address the need to reduce the potentially serious consequences of runway incursions, specifically those involving an aircraft on approach while another aircraft or vehicle is on the same runway. The FAROS provides a visual indication of runway occupancy status directly to landing pilots through the Flashing Precision Approach Path Indicator (FPAPI) FAA (2004). The MITRE Center for Advanced Aviation System Development (CAASD) conducted a simulation to examine Human Factors issues related to the proposed FPAPI implementation of FAROS.

Method

Experimental Task

Pilots were required to fly several approaches using both a standard PAPI and the new FPAPI system. They used the two PAPI systems to maintain the proper glide path and used the visual depiction of the runway to align themselves laterally. To minimize training time, pilots flew with the autothrottle engaged and set to the proper final approach speed. Their task involved tracking inbound to the airport, completing a short checklist, flying a stable approach, communicating with ATC, and determining runway occupancy status. All approaches were flown to runway 18 Center (18C) at Memphis International Airport (MEM). There was a continuous wind field beginning at 3000 feet from 220 degrees at 20 knots and decreasing to 10 knots from 210 degrees at the airport surface. This wind field was used for all trials. The time of day simulated a dusk environment that was clear of clouds with some light haze.

Experimental Design

Each pilot flew two trial types, baseline and experimental. The baseline trials were similar to today's environment with a steady PAPI and pilots were required to visually scan the runway to determine its occupancy status. The experimental trials included a FPAPI system, which provided pilots with a visual indication of the occupancy status of the runway. There were 18 baseline and 18 experimental trials for a total of 36 trials per pilot. The trials were blocked and pilots flew one block of 18 trials, took a short break, and then flew the other block of 18 trials.

Within each block of 18 trials, there were 16 trials with intruding traffic on the runway and two trials that did not include traffic. The no-traffic trials were included to provide pilots an opportunity to land without any traffic. These no-traffic trials were randomly presented within the block of 18 trials.

During each approach, the intruding aircraft entered runway 18C from one of two different locations. Half of the intruders entered the runway near the approach end at taxiway "Charlie 8" (C8) and the other half entered the runway midfield at intersection "Delta" (D). Half of the intruders entered the runway, positioned themselves for takeoff, and then remained in position until the end of the trial requiring the cockpit to execute a go-around in order to avoid a runway incursion. The other half of the intruders entered the runway, positioned for takeoff, remained in position for a few seconds, and then began the takeoff roll, thus clearing the runway in time for the cockpit to land. The intruders crossed the hold short line at two different points while the cockpit was approaching the runway. Half of the intruders crossed the hold short line when the cockpit was about 5 nautical miles (nm) from the threshold and the other half crossed the hold short line when the cockpit was about 2.5 nm from the threshold. Each of these three factors was completely balanced across the participants and randomly presented (without replacement) throughout the 16 approaches, which contained traffic, yielding a 2 (intruder type) x 2 (intruder location) x 2 (incursion timing) factorial within-subjects design of intruder type. Each of these eight intruder types was replicated, generating a total of 16 legs with intruders along with two non-traffic legs for each participant. Furthermore, the 18 trials were presented twice to each pilot, once as a baseline trial (no FPAPI system) and once as an experimental trial (with the FPAPI system). The order of presentation for the two trial types (baseline and experimental) was counterbalanced.

A single failure trial (miss) was presented to each pilot on the last trial within both the baseline and the experimental conditions. The miss trial was always presented as the last trial during each block in order to maximize the opportunity for pilots to develop trust in the system. This trial simulated a “lost” intruder that wandered onto the runway environment without being cleared by ATC. Thus there was no ATC communication with the “lost” intruder. Furthermore, the FPAPI system did not detect the intruder entering the runway, due to a surveillance failure accordingly, neither the FPAPI system nor ATC detected the intruder and a missed detection resulted. This yielded two trials per participant for a total of 36 failure trials.

Simulation Environment

The cockpit was an enclosed, fixed based, mid-fidelity transport aircraft simulator (see Figure 1). It was configured as a generic twin-engine, large weight category, jet aircraft. It had an autothrottle system, which was used throughout the evaluation to control speed. The simulation included audio capabilities supporting aircraft environmental sounds (e.g., slipstream noise) and ATC communication. A side-stick controller was used for aircraft control. The center pedestal housed the throttle quadrant, flap handle, and speed brake lever. Twenty-one-inch touch-screen displays were located in front of the left and right seat positions and displayed the Primary Flight Display (PFD) instruments and navigation information. A nineteen-inch display occupied the center instrument panel and displayed engine and flap status information. These comprised the Electronic Flight Instrumentation System (EFIS) displays. Pilots used the Precision Approach Path Indicator (PAPI) and out-the-window (OTW) depiction of the runway in order to fly the approaches and navigate to the runway. The OTW visual scene driver gave pilots a 130-degree virtual representation of the outside world. For more detail on the MITRE CAASD Air Traffic Management (ATM) simulation facility see Oswald and Bone (2002).

Figure 1. *MITRE Air Traffic Management Lab Cockpit Simulator*

Participants

Eighteen pilots were recruited for the simulation. Nine were classified as General Aviation (GA) pilots and nine were classified as Airline Transport Pilots (ATP) based on their experience. The GA pilots all indicated that they primarily flew piston aircraft (total flight hours M = 2097, SD = 2729, and range 109-8900). The ATP pilots all indicated that they primarily flew turbine aircraft (total flight hours M = 8798, SD = 6954, and range 2300-23000). All pilots were current within the previous three months.

Procedure

Upon arrival, pilots read and signed the informed consent form, filled out a short demographics questionnaire, and were given the experimental instructions orally. They were told that the experiment involved runway status automation and that the first trials were for training and familiarization with the simulator. Following this brief description, pilots were given some oral instruction about flying the simulator then four practice approaches were flown. A fifth trial was then flown in which the FPAPI system was activated when the intruder entered the runway environment. This was the pilots' first exposure to the FPAPI system and was intended to capture a naïve response to the system and elicit discussion following the practice trials. Due to space constraints, the naïve trial and subjective questionnaire data will not be discussed here see Helleberg (2004) for details. Pilots were then given a brief written description of the FPAPI system relating to the function, system design, and pilot procedures. After the pilot read the FPAPI system description, the experimental and baseline trials followed in a counterbalanced order (i.e., half of the pilots flew the baseline trials first, the other half flew the experimental trials first) with a short break between blocks of trials.

Each trial began with the cockpit simulator aligned with the runway and on the glide slope with the autothrottle engaged and set to the final approach speed. The pilot was told to assume control of the aircraft and fly the approach using the simulator's side stick to track vertically and horizontally to the runway. Along the approach the pilot was required to complete a short checklist (gear down and extend final flaps) as well as determine the occupancy status of the runway prior to landing. A confederate ATC provided normal take off and landing clearances on the tower frequency and responded to any spontaneous requests from the participant pilot. The pilot was responsible for making all radio calls.

Results

Pilot Experience

Across the dependent variables, there were no significant performance differences or interactions with the independent variables between the GA pilots and airline pilot experience groups. Therefore, the data from the two groups were pooled for the following analyses.

Decision Making Land/Go-Around

Each of the eighteen pilots flew 36 experimental trials, in which they were required to determine the occupancy status of the runway and make a decision as to whether it was safe to land or they should execute a go-around. This yielded a total of 648 trials available for analysis (see Figure 2). Each pilot was presented with four trials in which there was no intruding traffic and a clear runway. Across these 72 trials all pilots completed the approach and landed.

The remaining 576 trials included intruding traffic. On half of these trials the intruders remained on the runway thus requiring the pilot to execute a go-around to avoid a runway incursion. These trials will be referred to as "go-around" trials. On the other half of the trials the intruders departed the runway in time for the pilot to land without causing a runway incursion. These trials will be referred to as "landable" trials. During the landable trials the intruder would lift off of the runway when the pilot's aircraft was between three-quarter and one nm from the threshold. The number of go-arounds and landings, as well as the distance from threshold when the go-around call was made, were recorded and served as the dependent variables.

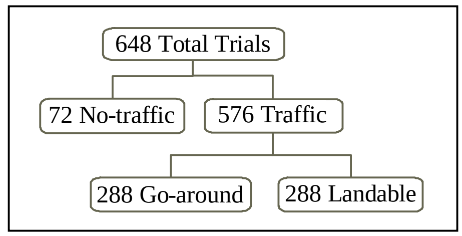


Figure 2. *Experimental Trials Across All Pilots*

Go-Around Trials Pilots initiated go-arounds on all 288 of the go-around trials. Therefore, regardless of whether the PAPI lights were flashing or not, none of the pilots landed on an occupied runway.

The distance from threshold when the go-around call was made was available on 269 of these trials (due to data collection errors, 19 of the trials did not have distance from threshold data). Across the 269 go-around trials, the FPAPI had no statistically significant effect on the distance from threshold when the go-around call was made (t (267) = 1.54, p ns). When the PAPI lights were steady (n = 135), the pilots made the go-around radio call at a mean distance from the threshold of 0.75 nm. When the PAPI lights were flashing (n = 134), the pilots made the go-around radio call at a mean distance from the threshold of 0.67 nm.

Landable Trials Pilots initiated go-arounds on 16 (6%) of the landable trials. Across these 16 trials, the FPAPI had no statistically significant effect on the number of go-around trials that occurred (χ^2 (1, N = 15) =0.25, p ns). Furthermore, the FPAPI had no statistically significant effect on the distance from threshold when the go-around call was made (t (13) = -0.33, p ns).

Across the 16 landable trials in which pilots elected to go-around, nine of those occurred when the PAPI lights were steady and seven occurred when the PAPI lights were flashing. Due to a data collection error, one of the FPAPI trials did not have distance from threshold data. Across the remaining 15 landable trials in which pilots elected to go-around, the mean distance from threshold when the pilots made the go-around radio call was 1.22 nm (n = 9) when the PAPI lights were steady and 1.39 nm (n = 6) when the PAPI lights were flashing.

Attention Allocation

During each approach, the pilots' were required to fly the aircraft and track their way to the runway surface using the PAPI for vertical guidance and the visual depiction of the runway for horizontal guidance. The pilots' ability to maintain the proper flight path was used as a measure of attention allocation.

Flight path Tracking The flight path tracking error calculation was conducted only on the portion of each trial that contained traffic. This was done to reduce any dilution of the errors during the portion at the beginning of each trial in which the lights were not flashing or, in the case of the steady PAPI trials, would not have been flashing. Root Mean Square (RMS) errors were calculated for both the lateral and vertical dimensions across the 576 traffic trials.

The FPAPI resulted in a statistically significant increase in lateral tracking errors ($F(1, 16) = 5.82$, $p < .03$). The lateral tracking errors increased from a mean of 60.0 feet during the steady PAPI trials to a mean of 66.8 feet during the FPAPI trials.

The FPAPI had no statistically significant effect on the pilots' vertical tracking performance ($F(1, 16) = 0.00$, p ns). The vertical tracking errors were similar between the steady PAPI trials (mean of 76.3 feet) and the FPAPI trials (mean of 76.1 feet). Figure 3 depicts the relationship between the state of the PAPI and flight path tracking performance.

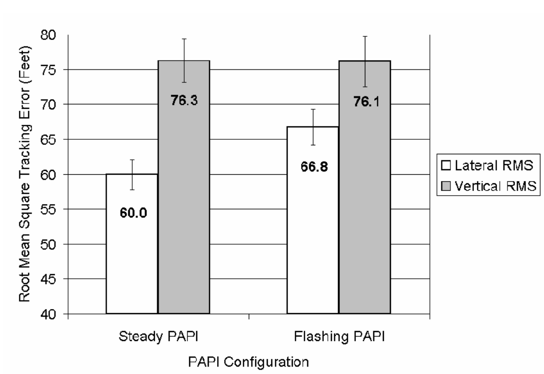


Figure 3. *Pilot Flight Path Tracking Performance Across Flashing Conditions*

ATC Communications

A confederate "air traffic controller" (ATC) was available to respond to pilot requests during the approaches. Whenever the pilot contacted ATC, the experimenter marked the data stream in order to derive the total number of calls as well as the distance from threshold when the call was made. The number of communications and distance from threshold results are described below. Eighteen pilots flew 36 trials each for a total of 648 trials available for analysis (see Figure 2). Each pilot was presented with four trials in which there was no intruding traffic and a clear runway. None of the pilots contacted ATC during the 72 trials that did not include intruding traffic.

The remaining 576 trials included intruding traffic. On the go-around trials pilots were required to make one radio call to report initiation of the go-around maneuver. On the landable trials, the pilot could complete the trial without contacting ATC. The number of trials in which pilots contacted ATC, as well as the distance from threshold when calls were made, were recorded as the dependent variables.

Go-Around Trials Pilots contacted ATC on all 288 of the go-around trials as well as the 16 additional landable trials in which pilots elected to go-around. This resulted in a total of 304 trials, which **required** one communication (i.e., notifying ATC of the go-around) for the following analysis.

Pilots completed 223 (73%) of the 304 go-around trials without making additional calls to ATC beyond the one required communication to ATC indicating that the pilot intended to execute a go-around. However, pilots made two or more communications on 81 (27%) of the 304 go-around trials. On eight (10%) of the 81 there were three communications. Across the 81 go-around trials with two or more communications, the FPAPI had no statistically significant effect on the number of trials in which pilots contacted ATC ($\chi^2(1, N = 80) = 0.01$, p ns).

The distance from threshold when the **initial** communication was made was available on 80 of these trials (due to data collection errors, one of the trials did not have distance from threshold data). Across the 80 go-around trials, the FPAPI had no statistically significant effect on the distance from threshold when the initial communication was made ($t(78) = -0.83$, p ns). When the PAPI lights were steady (n = 39), the pilots initially contacted ATC at a mean distance from the threshold of 1.56 nm. When the PAPI lights were flashing (n = 41), the pilots initially contacted ATC at a mean distance from the threshold of 1.46 nm.

Landable Trials There were a total of 288 landable trials and, on 272 (94%) of those trials, pilots completed the approach and landed. This resulted in 272 trials for the following analysis.

Pilots completed 223 (82%) of the 272 trials, in which pilots landed, without contacting ATC. However, pilots contacted ATC and made at least one communication on 49 (18%) of the 272 trials in which pilots landed. On three (6%) of the 49 trials there were two communications. Across the 49 trials, with at least one communication, the FPAPI had a marginally significant effect on the number of trials in which pilots contacted ATC (χ^2 (1, N = 48) = 3.45, p = .06). With the FPAPI there were 31 trials in which pilots contacted ATC and with the steady PAPI there were only 18 trials.

The distance from threshold when the **initial** communication was made was available on 42 of these trials (due to data collection errors, seven of the trials did not have distance from threshold data). Across the 42 landable trials, the FPAPI had no statistically significant effect on the distance from threshold when the initial communication was made (t (40) = -0.78, p ns). When the PAPI lights were steady (n = 13), the pilots initially contacted ATC at a mean distance from the threshold of 1.68 nm. When the PAPI lights were flashing (n = 29), the pilots initially contacted ATC at a mean distance from the threshold of 1.87 nm.

Complacency

Each pilot was presented with two "miss" trials, in which an unannounced intruder entered the runway environment requiring the pilot to execute a go-around in order to avoid a runway incursion. This intruder was always presented on the final trial of each block in order to maximize the opportunity for pilots to develop trust in the system. During both the steady and FPAPI trials, ATC would not clear the "miss" intruder to depart ahead of the pilot's aircraft (as had occurred during the previous 15 traffic trials). In addition during the FPAPI trials, the FPAPI system did **not** detect the intruder and the lights remained steady even though an intruder was located on the runway. The goal was to build the pilots' **expectation** that the FPAPI system would provide accurate information (across the preceding 17 trials) and then **surprise** the pilots with a system failure. However, this yielded a limited number of trials for analysis and accordingly the following results should be considered preliminary.

Go-Around Decision and Communications All 18 pilots detected both of the unannounced intruders regardless of whether the preceding 17 trials had been with the steady or FPAPI. The distance from threshold when the go-around call was made was available on 34 of these trials (due to data collection errors, two of the trials did not have distance from threshold data). Across the 34 miss trials, the FPAPI **expectation** had no statistically significant effect on the distance from threshold when the go-around call was made (t (32) = 1.20, p ns). When the pilots **expected** the PAPI lights to remain steady (n = 17), they made the go-around radio call at a mean distance from the threshold of 0.73 nm. When the pilots **expected** the PAPI lights to flash (n = 17), they made the go-around radio call at a mean distance from the threshold of 0.52 nm.

Additional Communications If the pilot contacted the confederate ATC during the miss trial and inquired about the runway status, the controller indicated that he could not see anyone on the runway, thus requiring the pilot to make his or her own determination of whether or not the runway was occupied. This frequently resulted in multiple calls to ATC. On 13 (36%) of the 36 miss trials, pilots made a single call to advise ATC that they were initiating a go-around. However, pilots contacted ATC two or more times on the remaining 23 (64%) miss trials. On four (17%) of the 23, there were three communications. Across the 23 trials with two or more communications, the flashing PAPI **expectation** had no statistically significant effect on the number of trials in which pilots contacted ATC (χ^2 (1, N = 22) = 0.39, p ns).

Furthermore, across the miss trials, which had multiple calls to ATC, the flashing PAPI **expectation** had no statistically significant effect on the distance from threshold when the **initial** call was made (t (21) = 0.81, p ns). When the pilots **expected** the PAPI lights to remain steady (n = 10), they made the initial call at a mean distance from the threshold of 1.56 nm. When the pilots **expected** the PAPI lights to flash (n = 13), they made the initial call at a mean distance from the threshold of 1.37 nm.

Discussion

This simulation was designed to address a set of Human Factors issues related to the proposed FAROS using the FPAPI system. The primary purpose of the simulation was to examine the most critical issues that could not be safely tested during an Operational Evaluation (OpEval). A secondary purpose was to collect some preliminary data related to several operational issues. However, due to the nature of simulation, these operational issues cannot be completely resolved and the data reported here should be combined with operational testing data.

.Pilot Experience

The pilots recruited for this simulation covered a wide range of experience levels. However, the results did not show any statistically significant differences in their performance during the simulation.

Land/Go-Around Decisions

None of the pilots landed on an occupied runway when the FPAPI was in use. However, none of the pilots landed on an occupied runway when the steady PAPI was in use either. This indicates that all 18 pilots visually verified the validity of the PAPI alert when it was flashing and also visually scanned the runway for traffic when the PAPI was steady.

There was no increase in go-arounds when pilots were flying with the FPAPI. Furthermore, there was no statistically significant difference in the distance from the threshold when pilots initiated their go-arounds due to the FPAPI. The data did not suggest that pilots were initiating go-arounds based on the FPAPI alone. Pilots tended to notice the lights flashing then shift their attention to scanning the runway for traffic as they neared the threshold. This suggests that pilots were using the status information provided by the FPAPI appropriately and FPAPI is unlikely to lead to an increase in unnecessary go-arounds.

Attention Allocation

The data revealed a statistically significant increase in lateral tracking errors associated with the FPAPI. However, there was not a corresponding increase in vertical tracking errors. One potential explanation for this result is that pilots may have focused their attention on the vertical tracking due to the attention capturing effect of the FPAPI. This may have led pilots to neglect their lateral tracking performance and concentrate on the vertical axis. However, the amount of lateral deviation was relatively small and may not be operationally significant, but should be considered when making the decision to move forward with an OpEval.

ATC Communications

The number of trials in which pilots contacted ATC and the distance from the threshold at the time of the call were used as objective measures of pilots' communications.

The data did not show any statistically significant increase in the number of trials, which contained communications due to the FPAPI. Also, the FPAPI had no statistically significant effect on the distance from the threshold when communications were initiated. However across the trials in which the intruder departed (landable), there was a trend suggesting that the FPAPI led to more trials with ATC communications. This suggests that a FPAPI could increase the number of ATC communications, however, it is possible that as pilots gain experience with FPAPI the number of ATC calls may decrease.

Complacency

All pilots detected both of the unannounced intruders regardless of whether the preceding trials had been with the steady or FPAPI. Furthermore, there was no statistically significant difference in the distance from threshold when the go-around call was made regardless of whether the preceding trials had the flashing or steady PAPI. Therefore, during the simulation the pilots did not show any evidence of complacency.

Acknowledgements

The author wishes to acknowledge the invaluable contributions of Ralph Lyons for help developing the simulation scenarios and running the participants, Scott Mills for data reduction and analysis support, and the software development team from the MITRE CAASD ATM Lab for their programming assistance.

References

Federal Aviation Administration (2004) Aeronautical Information Manual Washington, DC: Federal Aviation Administration.

Helleberg, J. R. (2004). Flashing Precision Approach Path Indicator Modeling and Simulation Final Report. (MITRE Technical Report MTR-04w0000051). The MITRE Corporation Center for Advanced Aviation System Development, McLean, VA.

National Transportation Safety Board (1991). Aircraft accident report: Runway collision of USAir Flight 1493, Boeing 737 and Skywest Flight 5569 Fairchild Metroliner, Los Angeles International Airport, Los Angeles, California, February 1, 1991 (NTSB/AAR-91/08). Washington, DC: Author.

Oswald, A. and Bone, R. S. (2002). Integrated Terminal Area and Flight Simulation Capabilities, Proceedings of the AIAA Modeling and Simulation Technologies Conference, Monterey, California 2002

EVIDENCE AGAINST CREW RESOURCE MANAGEMENT AS A COGNITIVE SKILL

Stacey M. L. Hendrickson
University of New Mexico
Albuquerque, New Mexico

David L. Trumpower
Marshall University
Huntington, West Virginia

Timothy E. Goldsmith
University of New Mexico
Albuquerque, New Mexico

Peder J. Johnson
University of New Mexico
Albuquerque, New Mexico

In recent years, airlines have begun to train and assess crew resource management (CRM) tasks similarly to technical tasks. However, in order for individual CRM categories (e.g., workload management, communication, situation awareness, etc.) to be viewed as skills, performance on a particular CRM category should transfer to different situations. In this study, we examined how well CRM behaviors generalized across different flight contexts. We analyzed pilot performance from five line oriented evaluations (LOEs). The LOEs were divided into phases of flight and many different behaviors were graded within each LOE, some of which were previously classified as belonging to a particular CRM category (e.g., workload management). A series of regression analyses showed that less than 1% of the total variance in grades was due to CRM categories; in contrast phase of flight accounted for roughly 10% of the total variance in grades. Thus, pilots performed more consistently within a phase of flight (regardless of CRM task category) than within a specific CRM category. We discuss several caveats and limitations associated with these findings. However, the findings do question the idea that CRM performance is a skill. One implication of these results is that pilot training may be more effectively focused around contexts rather than around specific CRM task categories.

Introduction

According to Welford (1976), a skill is defined by a high level of performance on a task that is achieved through training, is relatively permanent, and generalizes across similar situations. Training that aims to foster skill acquisition assumes that skills will generalize to contexts outside of training. Clearly this assumption is warranted in a broad sense; students do indeed gain expertise and go on to perform well in novel, real-world situations. However, questions remain about what kinds of task performances are best viewed as skills and what are the best methods for training and assessing these performances. These questions appear to be of particular relevance to the aviation community with respect to crew resource management (CRM). In recent years airlines have begun to train and assess CRM task performance in much the same way as technical skills. In this study, we investigated to what extent performance on CRM tasks such as situation awareness, decision-making, and workload management could be viewed as skills.

A related question is how effective is CRM training. In a recent review of its effectiveness, Salas et al. (in press) found mixed results, particularly when the assessment of CRM effectiveness was focused on behavioral outcomes. Other researchers have questioned CRM's effectiveness and validity too (Wiener, Kanki, & Helmreich, 1993). Perhaps CRM training has not been found to be more effective because the knowledge and behaviors being trained are in fact not skills, at least not in the traditional sense. An alternative and competing viewpoint is that CRM performance is contextually specific; performance on CRM-type tasks is more of a function of specific flight situations than the category of CRM behavior. If true, this would have important implications for training and assessing CRM.

Rather than examining the effectiveness of CRM training per se, we sought evidence that CRM performance behaved as an enduring trait across varied situations. More specifically, we investigated the construct validity of CRM categories by examining to what extent a pilot's performance in a single CRM category (e.g., decision making) was similar across two or more samples of this performance. If we know how a pilot performs on a particular CRM skill in one context, then we should be able to predict his performance for that same type of task in a different context. Conversely, the categories should also display discriminant validity in that we should be able to discriminate between the crew's performance on different CRM categories. We applied the classic psychometric paradigm of multi-trait multi-method (Campbell & Fiske, 1959) for investigating traits and situations to assess CRM skills of pilots. In the current study, we analyzed performance from a set of pilots who were evaluated along various CRM tasks that occurred in different contexts, in this case, phases of flight. If performance on CRM task categories is indeed skill-like, then we would expect to find higher similarity in performance of pilots within a CRM category than within a context.

Method

The basic data set we analyzed consisted of pilot performance data from 348 crews performing five LOEs under continuing qualification. Each crew was evaluated by an instructor/evaluator (IE) on the performance of 72 observable behaviors (OBs) over the course of an entire flight. Each LOE was comprised of 12 event sets (ES), each associated with a phase of flight (e.g., take-off, cruise, landing). Each crew was assessed on a 5-point grade scale by one of 20 IEs in one of the five LOEs.

Thirty-five of the OBs were intended to measure CRM performance (divided into five categories), and the other 37 OBs measured technical skills (divided into four categories). CRM categories were represented with OBs that might focus on the crew's communication (e.g. "The crewmembers state their ideas, opinions, and/or recommendations") or, as another example, they might address the crew's situational awareness (e.g. "The crew maintains shared level of situational awareness during precision approach").

Results

The grade distribution for all grades received by the crews are displayed in Table 1. Also shown is the breakdown for the grade distribution for CRM OBs and technical OBs. In all instances, the grade distribution was skewed so that the grades given were a majority of passing (greater than 2).

Table 1. *Proportion of grades received by flight crews*

	1 (unsatisfactory)	2	3	4	5 (excellent)	Mean (Std)
All OBs	<1	2	20	49	28	4.03 (.74)
CRM OBs	<1	1	24	47	28	4.02 (.74)
Technical OBs	<1	2	20	50	28	4.05 (.74)

Several multiple regression analyses were performed in order to assess the influence of the factors of specific CRM skill, context, and general skill on OB grades. Separate analyses were performed on CRM and technical grades. The dependent variable for each of these regression analyses was a crews' OB grade. Thus, each case corresponded to a single crew's grade on a single OB. For the CRM analysis, any given crew was represented by 35 cases since there were 35 CRM OBs in an LOE. With data from 348 crews, there were over 12,000 cases.

In the first analysis, we created predictor variables that reflected a crew's general skill, their performance on a particular skill category, and their performance in a given context. The three predictor variables were constructed as follows. First, for a given case, a mean grade was calculated for that particular crew's grades on all OBs from *different ESs* and *different skill* categories. This predictor was called General Skill (GS) as it represented a crew's performance across many phases of flights and skills. Second, an average was calculated for the particular crew's grades on all OBs of the *same skill* and *different ESs*. This predictor reflects a crew's performance in a particular CRM skill, and is thus called Skill (S). Third, an average was calculated for a crew's grades on all OBs from *different skills* but from the *same ES*. As this predictor represents a crew's performance in a specific event set, it was deemed Context (C).

For an example of how these variables were computed, consider a case associated with crew number 1 and a CRM-1 OB in the first event set. The GS predictor score for this case would be the average of crew number 1's grades on all non-CRM-1 CRM OBs in the second through twelfth ESs, the S predictor score would be the average of crew number 1's grades on all CRM-1 OBs in the second through twelfth ESs, and the C score would be the average of that crew's grades on all non-CRM-1 OBs in the first ES. Notice that all three predictors are orthogonal to one another; none of a crew's grades are used in the computation of more than one of the predictors.

From the regression analyses, a table can be constructed of the simple and semi-partial correlations among the CRM grades and the set of predictor variables. Results of this analysis are shown in Table 2. Overall, the three predictors accounted for 53.5% of the variability in CRM OB ra-tings. Although all three of the effects contributed significantly to prediction, the context effect was the strongest predictor accounting for 9.8% unique variance, while the skill and general skill effects uniquely accounted for just .3% and .1% of variance, respectively.

If, however, we consider the general skill effect as a baseline of performance, sequential multiple regression analyses can be performed to determine if the skill and context effects add anything to prediction of CRM OB ratings. Doing so, the skill effect accounts for just .1% more variability than the general skill effect alone, while the context effect accounted for 9.6% more variability than the general skill alone. Taken together, these results provide support for events sets, but very little for CRM skills, as individual units of LOE analysis.

Table 2. Squared **Zero-order** and Semi-partial Correlations Between Predictors and CRM OBs.

	Controlling for:					
Predictor	S	C	G	S & C	S & G	C & G
Skill (S)	**.378**	.017	.001	-	-	.003
Context (C)	.156	**.517**	.096	-	.098	-
General Skill (G)	.059	.015	**.436**	.001	-	-

Note. All values listed in table, except for values in bold, represent a semi-partial correlation between the CRM OB grades received and the constructed predictor variable in which the variable(s) listed along the top row of the table has been partialled out. If only one variable is listed, than it was the only one controlled for. Bold values represent the zero-order correlation of the constructed predictor variable with the CRM OB grades. Squared multiple R = .535.

Similar analyses were conducted in which technical OB ratings were predicted from the same three measures of general skill, skill, and context. These results are shown in Table 3. Overall, the three predictors accounted for 50.2% of the variability in technical OB ratings. Again, all three of the effects contributed significantly to prediction with the context effect emerging as the strongest predictor, accounting for 7.0% unique variance, and the general skill effect emerging as the weakest predictor in the model, accounting for less than .1%. The skill effect, however, accounted for 2.1% unique variance, somewhat more than in the CRM analysis.

Considering the general effect as the baseline of performance, sequential multiple regression analyses were also performed on the technical OB data. The skill effect accounted for 1.9% more variability than the general skill effect alone, while the context effect accounted for 6.7% more variability than the general effect alone. These results provide slightly more evidence for the validity of technical skills than that for CRM skills.

Table 3. *Squared **Zero-order** and Semi-partial Correlations Between Predictors and Technical OBs.*

	Controlling for:					
Predictor	S	C	G	S & C	S & G	C & G
Skill (S)	**.404**	.044	.019	-	-	.021
Context (C)	.098	**.458**	.068	-	.070	-
General Skill (G)	.028	.023	**.413**	.000	-	-

Note. All values listed in table, except for values in bold, represent a semi-partial correlation between the technical OB grades received and the constructed predictor variable in which the variable(s) listed along the top row of the table has been partialled out. If only one variable is listed, than it was the only one controlled for. Bold values represent the zero-order correlation of the constructed predictor variable with the technical OB grades. Squared multiple R = .502.

Several other analyses confirmed that technical OBs have relatively more skill structure than CRM OBs. Separate multiple regression analyses were conducted in which either CRM or technical OB ratings were predicted from the same Context and General Skill predictor variables as in the previous analyses along with several “skill” predictors. For prediction of CRM OB ratings, for example, scores for the Context and General Skill predictors were computed as before. In addition, a separate average was computed for a given crew’s ratings on all OBs within each of the five CRM categories. Thus, CRM OB ratings were predicted from each of the five CRM categories, Context, and General Skill scores. Likewise, technical OB ratings were predicted from each of the four technical categories, Context and General Skill scores. It should also be noted that separate regression analyses were conducted on OBs within each CRM and technical skill. Thus, a total of nine regression analyses were run; one in which only CRM-1 OB ratings were predicted, one in which only CRM-2 OB ratings were predicted, and so on.

For the technical OBs, the same skill score turned out to be the best predictor other than Context. That is, the tech-1 average predicted tech-1 OB ratings better than tech-2, tech-3, or tech-4 averages. This was not true of the CRM OBs (see Table 4).

Table 4. *Squared Semi-partial Correlations Between Predictors (Specific Skills, Context (C), and General Skill (G)) of CRM and Technical OBs.*

Outcome Variable	Predictor Variables						
	C-1	C-2	C-3	C-4	C-5	C	G
CRM OB Ratings							
CRM-1	**.010**	.002	.000	.001	.001	.055	.001
CRM-2	.003	**.001**	.000	.002	.000	.097	.000
CRM-3	.000	.001	**.001**	.002	.005	.116	.002
CRM-4	.002	.003	.003	**.002**	.000	.104	.002
CRM-5	.001	.000	.004	.000	**.001**	.104	.002
	T-1	T-2	T-3	T-4		C	G
Technical OB Ratings							
Tech-1	**.044**	.001	.000	.000		.053	.001
Tech-2	.000	**.033**	.000	.000		.046	.000
Tech-3	.000	.000	**.002**	.000		.091	.000
Tech-4	.001	.000	.001	**.014**		.125	.001

Note. All values listed in table represent a semi-partial correlation between the CRM (designated at CRM-1 or C-1 through CRM-5 or C-5) or technical (designated at Tech-1 or T-1 through Tech-4 or T-4) OB grade received and the constructed predictor variable in which all other variables listed along the top row of the table have been partialled out.

Finally, a cluster analysis was performed on the technical OBs and the CRM OBs to determine if an underlying skill structure existed for either type of performance. Four groups emerged for the technical OBs consisting of the four skill categories. However, the cluster analysis on the CRM OBs failed to show similar rankings by categories. The CRM grades were most notably clustered around event sets, giving further proof of a context effect. The distribution of the technical grades also demonstrated this effect along with the category clustering.

Discussion

The basic finding from our study is that pilots' performance within CRM task categories lacks consistency. We found higher consistency of performance within flight contexts (i.e., phases of flight) than within CRM categories. These results question whether performance within standard CRM task categories (e.g., decision making, situation awareness) should be viewed as skills in the traditional sense of that term. These findings have implications for how CRM should be trained and assessed. However, before we discuss these issues, we wish to bring up some possible criticisms and cautions of the current findings.

First, in the present analyses a CRM category is operationally defined by the specific OBs used to assess performance within that category. Were the particular OBs used in the current study good measures of the CRM task performance they purported to measure? In defense of the OBs we can say that they were written by experienced evaluators from a major carrier with a history of assessing and training CRM. Hence, they are likely to be as good as any in the industry. Further, previous research has shown that behavioral markers are capable of assessing the skills and knowledge typically associated with CRM categories.

Second, perhaps the IEs were poor at discriminating among levels of performance associated with CRM tasks. Others have questioned whether IEs can grade CRM performance with the same accuracy as they do technical skills. Objective qualification standards (e.g., +- 10 deg heading difference) govern the grading of technical tasks, but are often lacking for CRM tasks. Admittedly, CRM by its very nature is more subjective. However, we have found that IEs' inter-rater reliability for grading CRM performance was as high as for grading technical performance in previous studies of training and calibration sessions (Goldsmith & Johnson, 2002). We applaud efforts to improve the reliability and validity of measures of CRM performance, but it is likely that the evaluations of human performance in the aviation industry is as good or better than any industry.

A third caveat is the low variability in the performance data. A high proportion (77%) of the grades were 3's, 4's and 5's. The effect sizes for correlation and regression analyses are mitigated by skewed distributions and low variance. Could the small CRM skill effects be due to restrictions on the distribution of grades? Perhaps, but arguing against this is the fact that we found with the same performance data context effects that were substantially higher than CRM skill effects.

Assuming the results from our study are valid, what can we claim about the psychological status of CRM and what are the implications for assessing and training it? First, it may be that CRM performance is a skill but that the traditional CRM categories used to evaluate it are incorrect. The division of CRM into decision making, planning, workload management, etc. may not reflect the categories that best differentiate true cognitive performance. One way of determining psychologically valid categories would be through cluster or factor analysis on large sets of performance data. What skill categories emerge from the empirical data? The cluster analysis we performed on the performance data in the present study resulted in a single CRM category.

A second possibility is that CRM is a skill but rather than composed of a set of subskill categories (e.g., decision making, situation awareness, etc.) it is best viewed as a unitary, general skill. This idea is supported by the results of the cluster analysis in the present study, and also by the fact that the category skill effect accounted for only 1.9% more variance in the grades than the general skill effect alone. These results suggest that the division of CRM into categories has little explanatory power. The implication is that pilots do vary on CRM performance, but rating them along distinct CRM subcategories does not help much in differentiating their performance. If true, then what particular CRM tasks are trained and assessed is of less importance than their receiving some CRM training.

Finally, CRM may not be a skill at all. Psychologists have long debated whether traits or situations best characterize human personality and performance (Epstein & O'Brien, 1985; Michel & Peake, 1982). Many have questioned the idea that people have enduring characteristics that manifest across the varied contexts of life. Rather our behavior is more a function of the particular situation we find ourselves in. This same idea seems to best explain the data in the current study on CRM performance. If true, then our training and assessing of CRM performance should focus more on sampling flight contexts than on CRM tasks.

References

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the mutitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Epstein, S., & O'Brien, E. J. (1985). The person-situation debate in historical and current perspective. *Psychological Bulletin, 98*, 513-537.

Goldsmith, T. E. & Johnson, P. J. (2002). Assessing and improving evaluation of aircrew performance, *International Journal of Aviation Psychology*, *12*, 223-240.

Michel, W. & Peake, P. K. (1982). Beyond déjà vu in the search for cross-situational consistency. *Psychological Review, 89*, 730-755.

Salas, E., Wilson, K. A., Burke, C. S., & Wightman, D. C. (in press). Does CRM training work? An update, extension, and some critical needs. *Human Factors.*

Wiener, E. L., Kanki. B. G., & Helmreich, R. L. (Eds.). (1993). *Cockpit resource management*. San Diego, CA: Academic.

Welford, A. T. (1976). *Skilled performance: Perceptual and motor skills.* Oxford, England: Scott & Foresman.

LOWERED SITUATION AWARENESS WHEN USING A HELMET-MOUNTED HEAD-UP DISPLAY (HUD) IN A SIMULATED ROTARY-WING AIRCRAFT

Chris M. Herdman, **Jo-Anne LeFevre**, **Jerzy Jarmasz**, **Kamilla Johannsdottir**, **Lisa Hagen** and **Dan Bleichman**
Aviation and Cognitive Engineering (ACE) Lab
Centre for Applied Cognitive Research
Carleton University
Ottawa, Canada

Head-up displays (HUDs) have been shown to facilitate pilot performance in specific tasks such as controlling flight path and altitude. However, results from a number of simulator-based studies suggest that HUDs may decrease pilot situation awareness (SA) in tasks that require continuous monitoring of information in the environment. In extreme cases, HUDs have lowered SA to the extent that pilots may fail to detect potentially critical discrete events in the environment. Most research on HUDs has used fixed-panel displays. The present research examined the impact of a helmet-mounted display (HMD) HUD on pilot SA.

Introduction

Head-up displays (HUDs) have been introduced into aircraft in an attempt to enable aircrew to maintain a head-up/eyes-out view of the external environment while presumably simultaneously monitoring critical flight and power instrumentation. HUDs consist of symbology that is either located in a transparent fixed-panel that is located in the forward head-up view of the pilot or superimposed on a helmet-mounted display (HMD).

The motivation behind HUD technology is obvious: there are many situations where aircrew may need to look outside the cockpit while also closely monitoring aircraft instrumentation. For example, in a search and rescue mission aircrew may be required to maximize scanning of the environment while frequently viewing instrumentation to maintain a specific flight path and/or to strictly adhere to a set of flight parameters. To this end, research have shown that HUDs can be effective in controlling flight path and altitude (e.g., see Fadden, Ververs, & Wickens, 2001; McCann & Foyle, 1995; Martin-Emerson & Wickens, 1997; Wickens & Long, 1995). However, simulator-based studies have also shown that pilots tend to *cognitively tunnel* their attention onto a HUD to the extent that there are performance decrements on tasks requiring continuous monitoring of information in the environment (Foyle, Stanford, & McCann, 1991; McCann, Foyle, & Johnston, 1993), and in extreme cases, even failure to detect potentially critical discrete events in the environment (Brickner, 1989; Fischer, Haines, & Price, 1980; Wickens & Long, 1995).

To the best of our knowledge, published research on cognitive tunnelling with HUDs has only involved fixed-panel displays. Therefore, it is unknown whether cognitive tunnelling occurs with HMD HUDs. Whether cognitive tunnelling occurs with HMD HUDs is an important issue because this technology is becoming increasingly prevalent in aircraft.

The present research examined cognitive tunnelling with a HMD HUD and specifically whether pilot situation awareness (SA) was adversely affected by the HUD. Highly-trained Canadian Forces helicopter pilots flew a CH146 Griffon flight simulator through a series of simplified route-recce missions while wearing a HMD. Two conditions were compared: HUD versus No-HUD. In the HUD condition, the HMD was equipped with HUD symbology showing primary flight, power, and navigation information. The HUD symbology was derived from the Canadian Forces CH-146 Griffon helicopter Night Vision Goggle (NVG) HUD. In the No-HUD condition, the HMD was not equipped with HUD symbology. Instead, pilots were required to look under the HMD to read the head-down instrument panel. The No-HUD condition is similar to that typically experienced by CH-146 pilots using NVGs: when NVGs are not equipped with a HUD, pilots must look under the goggles to foveate the instrument panel.

Each pilot flew a series of the simplified route-recce missions. On each mission, pilots were instructed to provide reports (sitreps) of vehicle activity in the external environment (in the air or on the ground). Each mission was populated with a variety of ground entities (tanks, downed aircraft) and airborne entities (helicopter and fixed wing aircraft) that were placed to be close to the pilot's flight path. The entities varied in visibility, but all were visible for a minimum of 2 to 3 seconds. The primary measure of pilot SA was the number of critical entities in the external environment that the pilots were reported.

Methods

Participants. Eight male Canadian Forces CH146 Griffon pilots participated in this experiment. They ranged in age from 37 to 50 years and had between 10 and 29 years experience, with 1800 – 4800 hours total flight time and 780 – 1200 total hours in the CH146 Griffon. None had prior experience using either fixed panel or HMD HUDs. Thus, all were seasoned pilots but novice HUD users.

Design. A one-way repeated-measures deign was used with Condition as the only factor (HUD vs. No-HUD). Condition was counterbalanced across participants such that half received the HUD condition first and the no-HUD condition second. This order was reversed for the other half of the participants.

Flight Simulator. The experiment was conducted on the Networked Tactical Simulator (NTS) located in the Aviation and Cognitive Engineering (ACE) Lab at Carleton University. The NTS was configured to represent the cockpit, flight characteristics and general capabilities of the CH146 Griffon helicopter.

The NTS provides an immersive 180^0 (H) x 40^0 (V) out-the-window scene using three 8'x6' screens onto which image are presented via three data projectors. The NTS includes an emulated CH146 communications system and central display unit (CDU). The CDU provides a range of core functions to the CH146 aircrew.

For the purpose of this study, the simulator was equipped with a set of virtual reality goggles which served as the HMD. The goggles provided VGA resolution and a 37^0 (H) FOV which is similar to that of the NVGs used by CH146 pilots.

The HMD goggle was connected to an Intersense IS-900 Virtual Workbench head tracking system. The IS-900 is a 6 degrees-of-freedom tracker, tracking both position and angular changes (X, Y, Z, Heading, Pitch and Roll). The IS-900 provides position resolution of 1.5mm in position and an angular resolution of 0.05 degrees. It is jitter-free with a position stability of 4 mm and angular stability of 0.2, 0.4 RMS.

HMD HUD. The NVG HUD developed for the CH146 Griffon was modified and used in this study. The HUD was superimposed onto the external scene on the HMD. The HUD included the following symbology: heading tape, radar and barometric altitude, engine torque, slip ball, attitude indicator, and radio channel selection.

Terrain database. Pilots flew missions within a 10 km by 10 km terrain database of the Canadian Forces Base in Gagetown, NB. The terrain database contained a number of fixed, pre-determined geographical features (a river, hills, forest) and man-made elements (barracks, various military installations, roads, and the flight base). Various entities, both moving and stationary, were added to the terrain database to create the mission scenarios. Some of the entities were fixed navigation landmarks, which allowed pilots to follow pre-determined flight paths as instructed by the experimenters.

Targets entities used for assessing pilot SA. A number of entities were included in the simulation to provide pilots with objects to observe during their missions. These included two moving formations of three armored ground cars, three stationary pieces of artillery (Howitzer guns), four grounded CH-149 Cormorant helicopters, one wrecked CH-149, two CH-149s flying in small loop formations, two hovering CH-146 Griffon helicopters, one formation of four CF-18s flying in a wide formation across a large portion of the terrain, and one C-130 Hercules fixed-wing transport aircraft flying a slow, elongated loop pattern that cut across the whole width of the database terrain, roughly five kilometres from the southern edge of the terrain. All vehicles were placed so that they were on, or intersected, the paths pilots were flying in their missions. The CF-18s and the C-130 flew relatively slow and wide trajectories that intersected the pre-planned mission routes at fairly regular intervals. All entities were scaled to their normal size relative to the database.

Mission Scenarios. Two separate terrain databases were used in the experiment. Each database contained the same geographical features, buildings, waypoint markers, and entities (i.e., SA assessment task targets), and differed only in that the entities had different locations and trajectories in each database. Each pilot flew four missions in each terrain (one terrain on the first day of their participation, the other on the second day), for a total of eight missions per pilot.

Each mission consisted of flight legs (defined as a trajectory between two successive waypoint markers) arranged in a different order. The flight legs were sequenced such that (1) each waypoint was reached once per mission, (2) all the target entities were included on the path and distributed approximately equally between the legs of the mission, and (3) the legs constituted a continuous path starting and ending at the initial start point. Consequently, two successive legs could either be collinear or at an angle to each other (either 45º or 90º depending on whether both legs were on the edges of the database, or one was on a diagonal between a corner waypoint marker and the base). Thus, each mission was defined as a specific path visiting all eight waypoints, and was determined prior to starting the experiments.

Procedure. Testing of each pilot took place over two days. Each pilot flew 3 practice sessions on day-one before beginning the first of eight experimental sessions. For each mission, pilots were instructed to take-off from a base, and then head in the direction indicated by the experimenter. The pilot was directed to a waypoint, which was depicted in the database as a large white pylon: waypoint pylons were large enough to be seen regardless of flying altitude. Once the pilot had visually identified a waypoint, they were given a new heading for the next waypoint. They were then instructed to take an inside turn (if possible) around the waypoint and go to the new heading. Pilots were instructed to maintain an altitude of 200 feet above ground level (AGL) and an air speed of 80 knots throughout each mission.

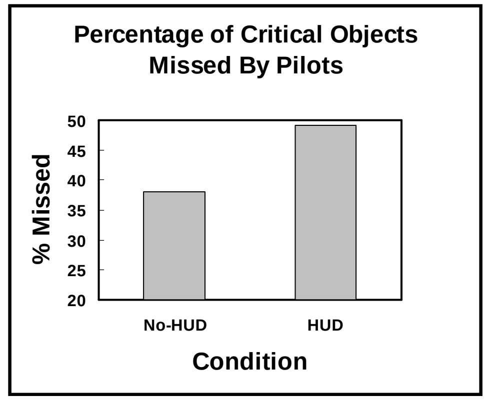
During the missions pilots interacted with the experimenter via the simulator radio communication system. Whenever the pilot spotted an object, they were to report it immediately. Experimenters kept track of the appearance of the entities, the number of entities (aircraft, ground vehicles, wrecks, etc.) reported by the pilot and the number of visible but missed entities. The experimenters had access to two computer screens located behind the pilot. One showed the actual scene the pilot saw at any given point in time, and the other showed the location of the ownship on a map of the database. This gave the experimenters full access to the scene being viewed by the pilot and allowed the experimenters to monitor pilot activities and the movement of the aircraft.

Each mission took approximately 20-minutes to complete. At the end of each mission, the pilot was asked to fill out a questionnaire that concerned their SA for objects in the environment.

Results

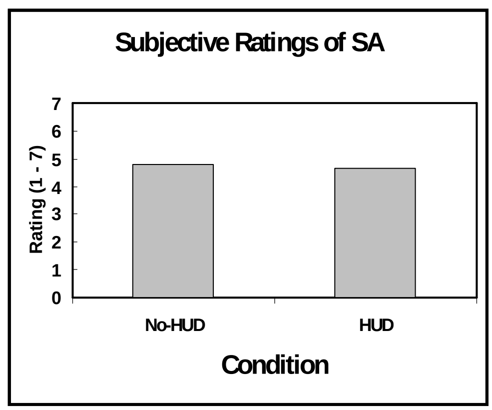
The results of this experiment are straightforward. Figure 1 shows the percent of objects missed by the pilots, summed across all missions. As shown in Figure 1, the HMD HUD lowered pilot SA in that pilots missed more objects in the HUD condition than in the No-HUD condition, $t(7) = 3.01$, $p < .05$.

Figure 1. *Situation Awareness Measure*



Post-experiment questionnaires showed that the pilots did not subjectively perceive their SA as being worse in the HUD than the No-HUD condition (see Figure 2). This suggests that the pilots were not subjectively aware of missing more objects in the HUD than in the No-HUD condition.

Figure 2. *Pilot SA Ratings*



Conclusion

The present study shows that an HMD HUD can lower pilot SA for objects in the external environment. The present findings are consistent with studies using fixed-panel HUDs in which it has been shown that pilots cognitively tunnel their attention onto a HUD resulting in decrements in performance on tasks requiring continuous monitoring of information in the external environment (Foyle, Stanford, & McCann, 1991; McCann, Foyle, & Johnston, 1993) or in some cases, failure to detect discrete events in the external environment (Brickner, 1989; Fischer, Haines, & Price, 1980; Wickens & Long, 1995).

Cognitive tunnelling has been linked to object-based visual attention whereby a HUD forms a perceptual object that is perceived and attended to separate from the external environment (Herdman et al., 2001; Jarmasz, Herdman & Johannsdottir, 2001; 2005; Martin-Emerson & Wickens, 1997; McCann & Foyle, 1995; Wickens & Long, 1995). On this view, cognitive tunneling represents the focusing of attention onto the HUD symbology at the expense of not attending to objects and events in the environment. The present research shows that cognitive tunnelling is not restricted to fixed-panel displays but can also occur with HMD HUDs. HMD HUDs are becoming common (e.g., in NVG systems). Thus, it is important to understand the parameters that affect cognitive tunnelling and to determine how to minimize the likelihood that cognitive tunnelling will occur.

References

Brickner, M. S. (1989). Apparent limitations of head-up displays and thermal imaging systems. In R. S. Jensen (Ed.), *Proceedings of the Fifth International Symposium on Aviation Psychology* (pp. 703-707). Columbus, OH: The Ohio State University.

Fadden, S., Ververs, P., & Wickens, C. (2001). Pathway HUDs: Are they viable? *Human-Factors*, 43, 173-193.

Fischer, E., Haines, R. F., & Price, T. A. (1980). Cognitive issues in head-up displays. *NASA Technical Paper 1711*, NASA Ames Research Center, Moffett Field, CA.

Foyle, D. C., Sanford, B., & McCann, R. S. (1991). Attentional issues in superimposed flight symbology. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 577-582). Columbus, OH: Ohio State University.

Herdman, C. M., Johannsdottir, K. R., Armstrong, J., Jarmasz, J., LeFevre, J., & Lichacz, F. (2001). Mixed-up but flyable: HMDs with aircraft- and head-referenced symbology. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics Volume Five: Aerospace and Transportation Systems* (pp. 73-82). Aldershot, UK: Ashgate.

Jarmasz, J., Herdman, C. M., & Johannsdottir, K. R. (2001). Object layers in HUDs: The role of motion in grouping symbology. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics Volume Five: Aerospace and Transportation Systems* (pp. 65-72). Aldershot, UK: Ashgate.

Jarmasz, J., Herdman, C. M., & Johannsdottir, K. R. (2005). Object-based attention and cognitive tunnelling. *Journal of Experimental Psychology: Applied.*

Martin-Emerson, R., & Wickens, C. D. (1997). Superimposion, symbology, visual attention, and the head-up display. *Human Factors*, *39,* 581-601.

McCann, R. S., & Foyle, D. C. (1995). Scene-linked symbology to improve situation awareness. *AGARD Conference Proceedings CP 575* (16-1 - 16-11). Brussels, Belgium.

McCann, R. S., Foyle, D. C., & Johnston, J. C. (1993). Attentional limitations with head-up displays. In R. S. Jensen (Ed.), *Proceedings of the 37th Annual Meeting of the Human Factors Society* (1345-1349). Columbus, OH: The Ohio State University.

Wickens, C. & Long, J. (1995). Object versus space-based models of visual attention: Implications for the design of head-up displays. *Journal of Experimental Psychology: Applied*, 1, 179-183.

Acknowledgements

This research was supported through NSERC and CRESTech grants to C. Herdman and by funding from DND, Canada. We thank the Canadian Forces CH146 Griffon pilots who volunteered for this project. We thank Major Scott Arbuthnott for sharing his insights and providing assistance with various aspects of the experiment, including mission scenario development.

Direct correspondence to Dr. Chris M. Herdman, Centre for Applied Cognitive Research, Carleton University, Ottawa, ON. Canada, K1S 5B6. Email correspondence: chris_herdman@carleton.ca

Improving Taxi Efficiency through Coordinated Runway Crossings

Becky L. Hooey
San Jose State University at NASA Ames Research Center
Moffett Field, CA 94035

The Coordinated Runway Crossing concept aims to improve the efficiency of airport surface operations by providing taxi clearances that contain a time or speed component. The goal is to enable pilots to arrive at, and cross, active runways without a delay. Eight commercial captains participated in a ninety-minute semi-structured interview that explored issues associated with coordinated runway crossings. The results of these interviews were used to generate preliminary information requirements, system requirements, and procedural requirements for a future coordinated runway crossing system.

Introduction

In current-day airport surface operations, the need to cross active runways while taxiing from gate to runway or vice versa often leads to extensive delays. One reason for these delays is that taxiing aircraft take third priority in runway usage (first priority is given to landing aircraft, second priority to departing aircraft). Another related reason is that air traffic control (ATC) tends to queue crossing aircraft on taxiways until there are sufficient numbers, and then cross all aircraft at once in a single crossing window between arrivals and departures. This approach to runway crossings is ultra-conservative, necessitated by the dynamic and uncertain nature of surface operations and the lack of information regarding predicted runway-occupancy times and predicted time-of-arrival of taxiing aircraft.

Efforts to increase aviation system capacity have focused a great deal of effort on improving arrival and departure efficiency. Ironically, these capacity-increasing initiatives, such as adding runways to existing airports and reducing the separation between aircraft on final approach, are likely to compound delays on the airport surface, and create new bottlenecks and problems (Cheng & Foyle, 2002).

As the aviation system expands to accept more traffic, airports are increasingly looking to alleviate the arrival and departure bottlenecks by adding runways, often adding one or more runways parallel to existing ones (Cheng, Sharma & Foyle, 2001). These additional parallel runways increase airport layout complexity, and displace the traffic bottleneck to surface operations. For example, Cheng and Foyle (2002) noted that when Dallas/Fort Worth (DFW) International Airport expanded from six runways to eight, the complexity of the airport configuration also increased. Adding runways resulted in some runways blocking the traffic between the gates and the outer runways causing more taxiway intersections and runway crossings to manage. The DFW Airport Development Plan (1997) proposed to address this via the addition of perimeter taxiways to route aircraft around the north and south ends of the runways. However, Cheng, Sharma, and Foyle (2001) note that this option is costly and results in increased taxi time and fuel burn.

An alternative proposal designed to increase capacity of the air transportation system is to reduce the spacing minima between landing aircraft. It has been proposed that speed cues from new on-board guidance systems will help enable precise spacing at the runway threshold (Barmore, Abbott, & Krishnamurthy, 2004) resulting in improved spacing consistency while eliminating excess spacing between landing aircraft, and increasing throughput. However, as arrival rates increase, there will be fewer opportunities for taxiing aircraft to cross active runways, and the length of available crossing windows may be shorter. Cheng, Sharma, and Foyle (2001) note that this is problematic given the current approach to queuing aircraft for a single runway crossing slot because an aircraft takes more than twice as long to cross the active runway if starting from a standstill as opposed to a continuous taxi (at 30 kts). Therefore, the number of aircraft that will be able to cross between arrivals may be reduced.

Coordinated Runway Crossing Concept. These 'solutions' to increase system throughput may have a significant negative impact on airport surface operations. In order to achieve the system-wide benefit that is expected, solutions must also be developed to enable efficient surface operations and runway crossings. If a controller could better predict gaps between arrivals and departures, and predict when a taxiing aircraft will arrive at the runway, aircraft could potentially be provided with a taxi clearance that enables them to cross the runway as they arrive, rather than waiting for a sufficient number of aircraft in a queue to cross at once. Cheng, Sharma, and Foyle (2001) concluded that this would not only reduce taxi time and delays due to hold-short operations, but also minimize taxi traffic back-ups, and ease the impact on the arrival and departures on the runway. Thus, in order to improve the efficiency of surface operations, the concept of coordinated runway crossings has been proposed which make use of new procedures, automation, and display technology to minimize or eliminate the need for aircraft to stop and wait to cross active runways.

There are many ways that a coordinated runway crossing system may be implemented, and each will have an impact on the task of the pilot, ATC, and the interaction between the two. The development of a human-centered system begins with an understanding of the current-day operations, and consideration of potential issues as perceived by operators within the system. This paper reports the findings from a series of interviews in which the objective was to initiate a dialogue with subject matter experts, in this case commercial pilots[1], to solicit their initial impression of the coordinated runway crossing concept and to identify issues that must be addressed in the subsequent research and development program. The results of these interviews were used to generate preliminary information requirements, system requirements, and procedural requirements.

Method

Participants

Eight commercial airline captains, representing five different U.S. airlines, participated in this study.

Procedure

Each pilot participated in a semi-structured interview that lasted approximately 1.5 hours. Each interview began with a discussion of the Captain's taxi experience, and issues faced during current-day taxi operations including factors that hinder airport efficiency and safety. Each captain described the airports they most frequently fly into and the typical runway crossing delays associated with each airport. Subsequently, a series of open-ended questions guided the discussion about procedures and technologies that could improve the efficiency of runway crossings.

The coordinated runway crossing system concept was introduced to the participants including a description of the intended system-wide efficiency gains expected from eliminating the need for aircraft to hold short of active runways. Two potential implementations (time-based, and speed-based taxi clearances) were described. For time-based clearances, two potential formats: Zulu[2] time and elapsed time were suggested. An example of a Zulu time command ("Cross Runway 22R at 22:13Z") and an elapsed time command ("Cross Runway 22R in 45 seconds") was presented. For speed-based clearances, pilots were told that clearances would contain a speed advisory such as "Taxi Alpha, Bravo, Charlie, maintain 16 kts". Such a clearance, if followed, would ensure a taxiing aircraft would be able to cross the active runway without delay. Focused questions were then asked regarding the implementation of each clearance type.

A transcript of the interviews was analyzed to identify sources of efficiencies in current-day surface operations. Preliminary information requirements, system requirements, and procedural requirements for coordinated runway crossing systems were generated. Given the semi-structured nature of the interview, the findings presented herein are qualitative in nature. They are presented with the intent to identify issues for consideration and guide subsequent research.

Results

Sources of Inefficiency in Current-Day Operations

Delays associated with crossing active runways were identified as the largest contributor to surface operation inefficiencies. Pilots cited delays of up to 20 minutes at some airports (including Dallas Fort Worth International, Phoenix Sky Harbor International, and Seattle-Tacoma International) and suggested that operations could be more efficient if the need to stop and start-up again could be eliminated. Pilots cited airport layouts as the largest source of inefficiency, particularly when the gates are on one side, and all traffic must cross active runways to get to their destination (gate or departure runway).

Pilots noted a lack of consistency among controllers attempts to maximize runway-crossing efficiency, primarily due to ATC workload and traffic loads. Under some conditions, controllers possess the ability to forecast future traffic patterns, and can therefore expedite traffic and minimize delays. For example, controllers may command longer taxi routes than the most direct route, if it actually minimizes the runway crossing delay and overall taxi time. However, this is not consistent, and controllers cannot accomplish it when it is needed the most, during peak traffic times.

[1] The importance of soliciting feedback from other stakeholders has not been overlooked. Similar interviews with ATC are currently being conducted.

[2] Zulu Time, also known as Coordinated Universal Time or Greenwich Mean Time, is used as the standard clock for international reference of time in communications, military, maritime and other activities that cross time zones.

This implies that some controllers are already trying to maximize efficiency at runway crossings by eliminating holds, but they do not have the information and the tools to do it consistently. More typical is the experience as described by one pilot:

> "At Phoenix and Dallas Fort Worth, ATC stacks up crossing aircraft until there are enough of them to break the arrival or departure stream. This leads to delays of 10 to 15 minutes."

Another reported source of operational inefficiency is the variability in response timeliness among pilots, which leaves ATC with uncertainty as to whether a pilot will respond to, and execute, the clearance promptly. If ATC issues a command and expects it to be carried out expeditiously, any delay in response could cause significant disruptions (e.g., could impact other traffic or cause a landing aircraft to initiate a go-around maneuver). Given the nature of the consequences, ATC must be more conservative in their commands and, if in doubt, issue a hold command rather than an expedite command. To illustrate this, one pilot reported:

> "Spacing is a lot closer for [X Airline] than other airlines because they are reliably fast and efficient".

Pilot Information Requirements

It is clear that both pilots and controllers in current-day operations are attempting to maximize efficiency, but lack the information needed to support coordinated runway crossings. Pilot information requirements are discussed below in terms of traffic, navigation and speed/time management.

Traffic. During the interviews, pilots suggested the need for improved sharing of information regarding traffic flow, aircraft sequencing, and runway use. For example, knowledge of upcoming breaks in the arrival stream would enable pilots to better gauge their taxi speed and be prepared to cross runways or take-off at the appropriate time.

> "The problem is not that I have to stop and wait to cross the runway, but that I have no information. If I know that I can't cross the runway for the next 5 minutes due to a heavy arrival stream, then I won't rush to get there".

The flow of relevant information about traffic and runway use can be improved in a number of ways. Most simply, this can be addressed procedurally, with ATC providing relevant information about traffic sequencing. Pilots suggested minor changes to ATC phraseology that could help pilots gauge the time urgency associated with a runway crossing command. Suggested examples included: "Cross after company '47", and "Traffic on 2 mile final, expedite crossing of Runway 26R". However, while these types of clearances are already used to a limited extent in current-day operations, it is problematic on a wide-scale because it adds to radio frequency congestion, and is often not possible at peak times. Another option for improving flow of traffic information is cockpit display technology that provides a real-time graphical depiction of traffic and runway occupancy to pilots. This shared awareness of runway traffic would lessen opportunities for errors and runway incursions, and could also help pilots cross-check ATC clearances.

Navigation. Pilots reported that their ability to accurately estimate and predict their time to arrive at a runway crossing point would be largely dependent on their familiarity with the airport. This suggests that flying into new and unfamiliar airports, or receiving a non-standard taxi route, could make complying with a time-based coordinated runway crossing command difficult for many pilots. On the other hand, those with routine and familiar routes stated that meeting a required runway crossing time would not be difficult.

> "For airports that I fly into, I know how many minutes it normally takes to get from runway to gate. I don't see this as a big problem."

This finding highlights the need for navigation displays that depict the airport layout, the location of the gates and runways, and the cleared taxi route. Such navigation displays have been developed (e.g., Hooey, Foyle & Andre, 2001; Theunissen, Rademaker, Jinkins & deHaag, 2002) and are under consideration for implementation by industry (Comstock et al., 2004). The wide-scale deployment of such displays should be considered a minimum requirement for the coordinated runway crossing concept.

Speed and Time Management. Pilots reported that their cockpits lack even the most basic speed and time management tools necessary to enable coordinated runway crossings. Specifically, many stated that their ground speed indicator is inaccurate at taxi speeds and this would make complying with a precise speed advisory unnecessarily difficult. Further, only those flying more modern aircraft are equipped with GPS-precision clocks. Therefore, complying with time commands in Zulu time formats would be difficult as it would required synchronization between pilots and ATC and confirmation of both actual and commanded times. Although most aircraft are equipped with stopwatches, using elapsed time creates other time synchronization problems, especially if pilots are slow to start their clocks in response to ATC time commands. Therefore, the inaccuracy of aircraft clocks, and the lack of synchronization among pilots and ATC, could contradict the precision required for closely spaced taxi maneuvers.

Minimal information requirements, then, include accurate speed indicators and synchronized clocks. Beyond these minimal requirements, further augmentations to cockpit displays will also be required. Displays that depict deviations between commanded and actual speeds in a graphical or status-at-a-glance format would aid pilots in the speed maintenance task. Displays that show both time elapsed and time remaining in an integrated fashion would allow pilots to better estimate their conformance to time-based commands. Pilots recommended a conformance monitoring system that would alert them when they are required to make a speed adjustment in order to attain their runway-crossing goal. The nature of the information that will be required, however, will be largely dependent on the required degree of precision with which the aircraft must arrive at the runway. The automation and display technology must be considered carefully to adequately support the required level of precision.

Summary of Preliminary Information Requirements
1) Traffic Management - Sequencing information - Location/intent of traffic 2) Navigation - Airport layout - Route depictions 3) Speed and Time Management - Accurate ground speed - GPS-precision clocks - Pilot-ATC synchronized timers

System Requirements

It is clear that a pilot's support for a coordinated runway crossing system would be dependent on the actual algorithms used to derive the speed or time commands. Factors that pilots determined must be considered in the development of a coordinated runway crossing system were grouped into four categories: Aircraft-specific, airport-specific, operating conditions, and traffic flow. Each is discussed below.

Aircraft-specific Characteristics. Speed or time-based advisories must be determined based on aircraft-specific minimum and maximum taxi speeds. The type of the aircraft will determine how quickly it can taxi and maneuver around turns. Airline policy, particularly policy regarding engine use during taxi, must also be considered. Some airlines require pilots to taxi on one engine, others taxi on two, at least until clear of all runways. This will influence how quickly an aircraft can taxi across a runway and prepare for take-off. Also, as technology is developed and adopted by airlines, the presence of technology on-board will influence an aircraft's ability to comply with time and speed commands. If cockpit display technology is gradually phased in, a system must be able to accommodate mixed-equipped fleets where some aircraft may be equipped with automation and display technologies to help them achieve their runway crossing time, but others are not. Unless it is clear which aircraft are equipped and able to comply, the result would be increased uncertainty for both ATC and pilots.

Airport-specific Characteristics. The runway crossing system must also be flexible enough to adapt to particularities at each airport. Characteristics that the system must consider include taxiway geometry, taxiway weight restrictions, and taxiways that are temporarily closed for maintenance. Also, many airports have unique characteristics for which the system must be adaptable. For example, Las Vegas McCarran (LAS) Airport has a long downhill taxi and requiring a pilot to ride the brakes to maintain a slow taxi speed could overheat the brakes creating a threat to safety in the event of an aborted take-off. New York's LaGuardia Airport (LGA) has some taxiways that limit taxi speed to 5 kts due to poor surface conditions. The ability for each airport to apply constraints based on their temporary and permanent taxiway circumstances will be required.

Operating Conditions. Pilots listed a number of operating conditions that limit or otherwise affect the speed at which a pilot can taxi. Specifically, the pilots advised that the system must be able to adapt speed and time commands to account for slower taxi speeds necessitated by poor visibility and surface friction conditions. Operational conditions that create the need for de-icing before take off must also be considered to ensure time/speed commands enable aircraft to taxi efficiently from the de-icing station to the departure runway and eliminate delays which will cause an aircraft to return to the deicing station.

Traffic Flow. Traffic flow is a large consideration in developing the time or speed algorithms. Clearly, a coordinated runway crossing system must include intelligence to allow for coordination among aircraft so that a following aircraft is not commanded to taxi faster than the lead aircraft. Similarly, aircraft cannot be sent along conflicting paths (a particular problem near gate alleyways), or at least their speeds must be adjusted to prevent conflict while still reaching their target runway crossing time.

Summary of Preliminary System Requirements
1) Aircraft Specific Characteristics - aircraft type - airline policy - equipage 2) Airport Specific Characteristics - taxiway geometry - taxiway weight restrictions - closed taxiways 3) Operating Conditions - visibility - surface friction - icing conditions 4) Traffic Flow - gate assignment - routing and speed conflicts - runway usage

Procedural Requirements

The success of a coordinated runway crossing system requires more than new technology and cockpit displays. It also requires the simultaneous development of new operational procedures. Pilots raised four procedural issues that merit consideration.

Contingency Plans. During the interviews, pilots emphasized the need for contingency plans that would accommodate circumstances in which a pilot cannot comply with a crossing command, and do so without heavily penalizing them with lengthy delays. The most important concern raised repeatedly by the pilots was that a coordinated runway crossing system could promote unnecessary rushing or a 'rush to comply mentality' in which crews rush through checklists and other duties in order to meet their runway crossing and departure times. This is particularly problematic given that ATC would not be aware of situations in the cockpit where the crew may be struggling with navigation or other cockpit duties. Pilots noted that it impossible to predict the time required to complete these tasks as it will depend on airline procedures, flight-crew experience, and cabin crew experience. Requiring pilots to maintain a specific speed or arrive at the runway at a specified time means the crew may arrive at the runway before they are prepared to take off because they have not completed checklists and safety items. It could also lead to the dangerous situation in which the first officer is removed from navigation tasks in an effort to attend to other tasks that must be completed.

Contingency plans could take the form of automatic adjustments to runway crossing slots based on taxi speed and/or speed conformance monitoring. A time-based clearance might take the form "Cross at 22:10Z, if unable expect next crossing at 22:35Z". This would allow pilots to assess their workload and ability to make the crossing, and at the same time be aware of the consequences of missing the window.

ATC and Pilot Interaction. Procedure development must also define phraseology for pilots to communicate to ATC in the event that they cannot make their cleared runway-crossing window. If ATC receives this information early then route modifications can be issued and the runway-crossing slot can be reassigned to another aircraft to make efficient use of the runway, thus maintaining the intended efficiency of the system. Similarly, there could be a need for ATC to cancel a runway crossing command, if for example, an aircraft is slower to land than expected, or an aircraft aborts take off. Effective means to communicate this information with standard phraseology must be developed.

Need for Positive ATC Control. Pilots emphasized the need for positive control at the runway crossings, (i.e., the need for ATC to verbally clear them to cross the runway, rather than simply provide a runway crossing window or time in the clearance). However, some pilots suggested this could take the form of datalink and display technology, not necessarily just the verbal commands over radio, as used today.

Conditions of Use. Procedures regarding when the system should be employed must also be developed. Pilots cautioned of 'system over-kill', suggesting that speed-based guidance should only be provided when relevant, otherwise pilots will ignore the advisories. If the airport traffic is light, the system may encourage a pilot to taxi at maximum speed, only to discover that the gate is blocked. These usage procedures must be generated in coordination with both the airlines and pilot unions.

Summary of Preliminary Procedural Requirements
1) Contingency plans 2) ATC-pilot interaction 3) Positive control of all runways 4) Conditions of use

Discussion

The need to cross active runways during taxi leads to highly inefficient operations. The dynamic and uncertain nature of surface operations, coupled with the lack of information regarding predicted runway-occupancy times and predicted time-of-arrival of taxiing aircraft, requires ATC to be overly conservative and queue aircraft to cross the active runway as a group by building a gap between arriving and departing aircraft. This causes delays, sometimes in excess of 20 minutes, for taxiing aircraft and makes inefficient use of runways.

Evidence was provided to suggest that both controllers and pilots already attempt to improve the efficiency of surface operations by approximating coordinated runway crossings as time and workload levels permit. They accomplish this in a number of ways such as issuing an expedited crossing clearance, or by issuing or requesting a taxi route that is longer in distance but circumvents the need to hold short of an active runway. However, it was noted that the pilots and controllers lack the information and tools to do this consistently, and are unable to do this under high traffic loads, when it is most needed.

It is proposed that the concept of coordinated runway crossings, if accompanied by supporting procedures, automation, and display technology, could potentially increase the efficiency of airport surface operations by reducing hold delays and improving runway usage. Pilots indicated that the proposed coordinated runway crossing concept could be valuable to handle the traffic congestion problem, particularly if traffic flow increases as is predicted over the next several years. As one pilot remarked:

> "… with plans to reduce vertical separation, and with more airlines moving to smaller aircraft, there will be a big crunch on the airport surface. Smart movement of aircraft on the ground will be critical."

Several pilots highlighted the potential value of the system to help standardize taxi speed and conformance. Many noted that ATC currently must manage a great amount of uncertainty with some pilots responding quickly and others slower to comply. This uncertainty requires larger separation between aircraft. Pilots suggested that the separation could be reduced, and still be safe, with a coordinated runway crossing system and displays which increase pilot-ATC shared awareness and integrate traffic and runway information. These displays could improve awareness of runway traffic, lessen opportunities for mistakes and runway incursions, and could help pilots cross-check ATC clearances.

However, despite the general approval of the pilots involved in these interviews, a large hurdle that must be overcome before the development of a coordinated runway crossing system is to ensure user acceptance on a wide scale. Not surprisingly, some pilots felt that dictating taxi speed could be perceived negatively by pilots and could be met with resistance. It is important that the system demonstrates the value or benefit to the pilot and provides a clear rationale for the speed requirements (i.e., taxi 20 kts to cross in front of landing aircraft or hold for 10 minutes).

For this concept to be successful, a human-centered approach will be required that involves participation from pilots, ground controllers, local controllers, ramp controllers, and airlines. The pilot interviews reported in this paper represent the first investigation of the coordinated runway crossing concept with subject matter experts. Similar investigations with other stakeholders in surface operations are planned, as are human-in-the-loop simulations to assess pilot conformance to speed- and time-based clearances.

References

Barmore, B., Abbot, T., Krishnamurthy, K. (2004). Airborne-managed spacing in multiple arrival streams. International Congress of the Aeronautical Sciences. (Aug 29 – Sept. 3) Yokohama, Japan.

Cheng, V. H. L., Sharma, V., & Foyle, D. C. (2001). A study of aircraft taxi performance for enhancing airport surface traffic control. IEEE Transactions on Intelligent Transportation Systems, 2(2), 39 - 54.

Cheng, V. H. L., & Foyle, D.C. (2002). Automation tools for enhancing ground-operation situation awareness and flow efficiency. AIAA Guidance, Navigation, and Control Conference, Paper AIAA 2002-4856.

Comstock, R., Corker, K., Endsley, M., French, G., Prinzell, L., Snow, M., Wickens, C., & Etherington. T. (2004). Human Factors Issues in Synthetic Vision Displays: Government, Academic, Military, and Industry Perspectives. Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society, 75-78, Santa Monica: HFES.

Dallas/Fort Worth International Airport Board (1997). Airport Development Plan Update—Dallas/Fort Worth International Airport.

Hooey, B. L., Foyle, D. C., & Andre, A. D. (2001). The design of aircraft cockpit displays for low-visibility taxi operations. In A. G. Gale (Ed.) Vision in Vehicles IX. Holland: Elsevier Science Publishers.

Theunissen, E., Rademaker, R.M., Jinkins, R.D., & de Haag, M. U. (2002). Design and Evaluation of Taxi Navigation Displays. Proceedings of the 21st Digital Avionics System Conference, Irvine, CA.

Acknowledgement

This research was funded by NASA's Airspace Operations Systems/HMP/HADMT and Virtual Airspace Modeling and Simulation programs, and supported by Dr. David Foyle of NASA Ames Research Center's Human-Centered Systems Laboratory.

EXPERIMENTAL ANALYSIS OF TASK PRIORITIZATION TRAINING FOR A GROUP OF UNIVERSITY FLIGHT TECHNOLOGY STUDENTS

Amy L. Hoover
Central Washington University
Ellensburg, Washington

Kenneth H. Funk II
Oregon State University
Corvallis, Oregon

The purpose of this study was to evaluate changes in task prioritization performance between pilots who participated in a CTM training course and those who did not. A pretest-posttest control group design with random assignment was used. Pilots enrolled in the Central Washington University Flight Technology Program flew pretest and posttest simulated flights on a Frasca FTD. During a two week period between pretest and posttest simulated flights pilots in the experimental group participated in a CTM training course and pilots in the control group did not. Comparison of pre- and posttest error rates shows the experimental group had a 54% decrease in task prioritization errors and the control group had a 9% increase in errors.

Introduction

Pilots routinely perform multiple, concurrent tasks and the ability to effectively prioritize them for attention is a critical flying skill. Although pilots clearly understand this and generally practice concurrent task management (CTM) well, there are many instances in which failure to properly prioritize tasks or otherwise manage them effectively has led to a potentially dangerous incident or even a fatal accident (Chou et al, 1996).

Short-term memory appears to be a major limiting factor in CTM performance, so it is not surprising that a computational aid to augment human memory facilitated CTM performance in a low-fidelity flight simulator experiment (Funk and Braune, 1999). But technological limitations and other practical considerations strongly suggest that other means of improving CTM be explored, notably the training of CTM skills, including that of prioritizing tasks.

Bishara and Funk (2002) developed and evaluated a short (two-hour) CTM training module for general aviation pilots. In a pretest-posttest control group experiment, participants who received CTM training showed improvement in prospective memory performance. But results relating to task prioritization, a more general subskill, were ambiguous. This may have been due to several factors, including the quality of the training material (not developed by qualified flight instructors), the low fidelity of the simulator (Microsoft Flight Simulator was used), a small sample size (12), and the heterogeneity of the participants (reflecting a wide range of experience and skill). Although CTM performance is a significant factor in flight safety, the trainability of CTM, until now, has been in question.

Objective

The objective of this study was to carefully develop and evaluate CTM task prioritization training in a higher fidelity experimental environment using a more homogeneous population of participants.

Method

A pretest-posttest control group design was used. All participants flew a one hour simulated instrument flight on a Frasca 141 FTD (pretest) then flew another simulated flight two weeks later (posttest). The experimental group participated in a CTM training course during the two week interim and the control group did not.

Participants

Twenty-seven pilots enrolled in the Central Washington University Flight Technology Program participated in the experiment. Participants were randomly assigned to either the experimental or control group. All pilots had logged previous instrument time on the FTD used in the experiment. Regression analysis showed no correlation between participants' total flight time, instrument time, stage of training, total FTD time, and Frasca 141 FTD time with regards to CTM performance on the pretest, indicating the two groups were equivalent.

Flight Training Device

Two identical Frasca 141 FTDs were used and were configured as normally aspirated single engine fixed gear aircraft. The Avionics package included audio panel with marker beacons, dual VHF communication and navigation radios, DME, ADF, and a Garmin GNS430 IFR enroute and approach certified GPS/comm. The FTDs recorded all primary flight data including aircraft heading, altitude, airspeed, power settings, and position.

Procedure

Pre- and posttest simulated flights were conducted in a line oriented flight training (LOFT) format. The LOFT placed pilots in a high workload environment in Seattle Class B airspace and included radar vectors as well as pilot navigation, two precision instrument approaches, a multistage missed approach, and a holding procedure. Pilots conducted the simulated flights as per the CWU Standard Operating Procedures (SOP) manual; all checklists, flow checks, and callouts were the same used in their normal flight training.

Certified instrument flight instructors (CFIIs) were trained to administer the LOFT which was scripted with respect to air traffic control (ATC) communications and procedures. Flights were observed and scored in real time and again from videotape by a second scorer. Video cameras recorded a wide angle view that included the entire instrument panel, engine controls, yoke, rudder pedals, and pilots' hands and feet.

Prioritization scheme used

A task prioritization scheme taught to pilots during primary and advanced flight training is the aviate, navigate, communicate (ANC) hierarchy (Chappell, 1998; FAA, 1999; Jeppesen, 2001, 2003a, 2003b; Kern, 1998; Kershner, 1998; Machado, 2001, 2003; Thom, 1991). For this study each task was defined based on pilot training manuals and literature as follows:

Aviate task: Included all items related to aircraft operation: airspeed, altitude, climb or descent rate, lift, thrust, and drag; e.g. primary aircraft control inputs (pitch, power, yaw, and roll), operation of lift and drag devices (flaps) and operation of primary engine systems.

Navigate task: Included items related to the current and future position of the aircraft, including vectors, course intercepting and tracking, identification of intersections and waypoints, and programming and operating the GPS and other navigation radios.

Communicate task: Included communications with ATC.

Definition of CTM Errors

Opportunities for twenty potential task prioritization errors were embedded at 14 challenge points throughout the one hour simulated flights. Challenge points were based on errors observed during a pilot study conducted prior to the experiments. Each challenge point provided an opportunity for the participant to divert his/her attention from a more important or more urgent task to a less urgent or less important task. Associated with each challenge point were specifications as to what actions would constitute which type of prioritization error. Types of prioritization errors included ignoring an aviate (flight control) task in order to navigate (aviate/navigate, 7 opportunities), aviate/communicate (7), navigate/communicate (5), and aviate/aviate (1) in which the pilot had to choose between two aviate tasks as to which was most critical to perform first.

Several of the challenge points were simply part of the LOFT scenario; they were embedded at a point where a pilot might make a task prioritization error and thus did not require any intervention. For example, challenge points were placed at locations in the flight scenario where there was potential for error if the pilot fixated on or became distracted by a navigate task at the expense of primary aviate tasks. Other challenge points required the CFII to act as ATC and call the pilot with information or instructions just before the pilot was leveling off or about to intercept course, or to cause a failure to a navigational facility or an aircraft system.

Performance criteria for determining if an error occurred was based on FAA-S-8081-4C Instrument Rating Practical Test Standards with respect to altitude, airspeed, heading, intercepting and tracking course, use of checklists, procedures, and ATC communications.

CTM Training Course

The training course followed standard practices and procedures common to the CWU training course outline (TCO) and university criteria for learning outcomes and assessment strategies. The course was taught by an FAA certified CFII and CWU flight technology professor. It consisted of two sessions 7 days apart that included reading, self-study, cooperative learning activities, guided discussion, and a reflective homework assignment. The course also emphasized procedural discipline with respect to task prioritization, including proper use of checklists, standard operating procedures, mnemonic aiding devices, situational awareness, and cockpit flow checks.

The first learning session consisted of a class discussion of selected materials related to aviation human factors, aeronautical decision making, situational awareness, workload management, and concurrent task management. Participants had prior knowledge of all those concepts from previous coursework and studies, thus the training did not introduce any new concepts but rather emphasized task prioritization as an important element of human factors and aeronautical decision making. Participants analyzed accident and incidents taken from the NTSB and NASA databases with respect to CTM errors and participated in class discussions of those data.

During the time between sessions participants were asked to reflect on at least one of their normal flights with respect to CTM concepts and how their awareness influenced their in-flight decision making. Students reflected in writing as well as through a verbal debriefing.

The second class session included an activity in which participants acted out role-playing scenarios designed to give insight into their reactions and behavior in the cockpit when confronted with CTM challenges. They also participated in a class discussion of strategies to improve pilot task prioritization performance and a guided discussion of the outcomes. A short quiz was given at the end of the second session to evaluate each pilot's progress and identify areas of improvement.

Results

CTM error data were recorded as a frequency distribution of raw scores and converted to a ratio score (number of errors: total number possible) for comparison. Table 1 and Figure 1 present CTM error scores for experimental and control groups. The control group showed a 9% increase in total CTM errors, and the experimental group showed a 54% decrease in total errors.

Table 1. *Task prioritization error rates for each group. Mean scores are shown with standard deviation in parentheses.*

Group	Pretest	Posttest
Experimental	0.24 (0.12)	0.11 (0.08)
control	0.23 (0.15)	0.25 (0.10)

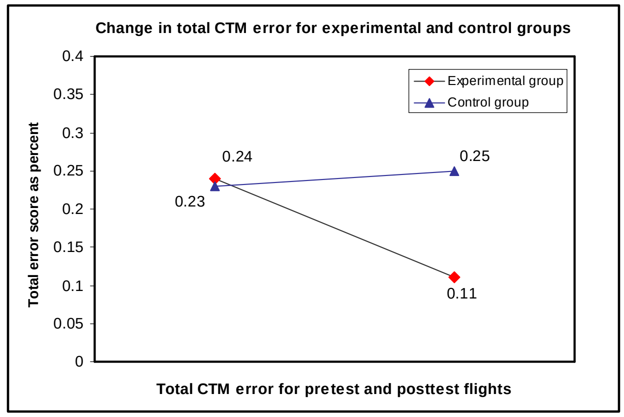


Figure 1. *Graph showing the change in total CTM error scores for each group expressed as a percent of total possible errors.*

There were 14 in the experimental group and 13 in the control group. An F-test for homoscedasticity found the samples had equivalent variance and K-S test and Q-Q plot showed they were normally distributed, so an independent samples t-test was used to compare the two groups. Because data showed a posttest reduction in CTM errors for the experimental group compared to the control group a one-tailed test was used yielding a t = 2.67 at p = 0.007 (Table 2).

Table 2. *Independent samples t-test*

t-test for equality of means				
		t (t critical)	*df*	Significance (1-tailed)
Posttest-pretest differen ce	Equivalent variances assumed	2.67 (1.71)	25	0.007
	Equivalent variances not assumed	2.68 (1.71)	24.3	0.006

Discussion

Results show the control group made the same or more prioritization errors overall in the posttest flight compared to the pretest; individual pilots showed an increase, a decrease, or no change in errors. Such a distribution would be expected from randomly sampling a group of pilots during two discreet flights. If there were no effect from the CTM training course then the experimental group should show a similar distribution of pretest and posttest scores. However, the experimental group had a much larger decrease in total CTM errors between pretest and posttest flights compared to the control group.

It seems reasonable that any well designed training course would show some effect during the short term, but a major question that arises is; how long will it take that effect to disappear, or to drop below acceptable performance standards? The answers to those questions would need to be assessed by testing the same participants at a later date, as well as controlling for effects of extraneous variables that might affect their performance.

The amount of time between the pretest and posttest simulated flights (2 weeks) represented a trade-off between internal and external validity. The time period was kept short enough to reduce history effects, but that did not allow the study to comment on longer term effects of the training. For a longer period of time, control for extraneous variables, including further training in human factors and additional flight experience, might be difficult. However, pretest data indicated no correlation between this particular group of participants' total flight time, instrument time, or FTD time and their CTM performance, so controlling for the influence of such extraneous variables might be a reasonable possibility.

A related question is whether or not any learning actually took place; pilots who received training showed a decrease in CTM errors and an improvement in performance over a two week period of time, but it is not known from this study whether they actually retained the new information or learned new behaviors that will endure.

Since all pilots in the experiment had previously studied concepts of prioritization and task management during their regular flight training, it is possible that the reduction in CTM errors by the experimental group might represent a sensitization effect; the only difference between the two groups could have been that the experimental group was focused on those concepts during the short term and did not actually code the information into their long term memory.

The issue of whether learning occurred is a critical one and also difficult to resolve because a teacher or instructor often does not have the ability to evaluate students after they leave the learning environment. More follow-up studies are needed to comment on the long term effects of the training. Additionally, a qualitative response from participants at some future time might also reflect on whether or not they felt learning occurred.

Pilots in the experimental group who showed the greatest reduction in CTM error scores were the ones that originally made the most errors. Thus it could be that the reduction in errors might simply represent a regression toward the mean for those pilots. However, the fact that several pilots in the control group also scored a large number of errors in the pretest without a corresponding reduction in errors for the posttest indicates that regression was probably not the cause for that trend in the experimental group.

What the data does suggest is that pilots who performed the worst seemed to benefit more from the training than those who initially made a low number of errors. Alternatively, pilots who made only one or two errors in the pretest and posttest were not able to be evaluated with respect to a training effect since there were only a fixed number of challenge points and it was not possible to show a large improvement in error scores for those pilots.

One error that more than half the pilots in both groups made involved a missed approach procedure (MAP) that called for the pilot to climb via the localizer course to 2000 feet, then to identify a specific intersection as the point to commence climb to 5000 feet while continuing to track the localizer course. Many of the pilots became fixated on the task of either programming the GPS for the waypoint or tuning and identifying the VOR to identify the cross radial for the intersection and either strayed off course, deviated from altitude, or both, while attempting to identify the fix. In several cases the video tapes showed pilots were not even looking at their flight instruments while operating the GPS unit. A few pilots were off altitude by as much as 500 feet and off the localizer course by a full needle deflection as a result of their fixation on the navigate task.

The issue of fixation has become an area of great concern in the flight training industry in recent years; over the past 5 years general aviation cockpits have incorporated more sophisticated IFR certified GPS units, and in the past 3 years flat panel primary flight displays (PFDs) and multifunction displays (MFDs) have been installed in training aircraft.

Wilson (1998) found that as the level of sophisticated instruments and automation increases on airline flight decks the potential for CTM errors also increases. Also, in a more general meta-study of airline flight deck human factors issues, Funk et al (1999) found the attentional demands of automation to be problematic. It is likely that the same potential exists for increased sophistication in general aviation cockpits, including training aircraft. Pilots who preprogrammed the GPS while still on the ground at did not make that fixation error.

Conclusions and Recommendations

Experimental analysis showed that the group of university flight students who participated in the CTM training course improved their task prioritization performance over a two week period of time. The decrease in task prioritization errors for pilots in the experimental group did not seem to be a result of regression toward the mean. It is not certain whether that performance increase had a longer lasting effect.

Pilots who did not participate in the CTM short course did not markedly improve in their prioritization performance; they either showed an increase, decrease, or no change in performance.

One particular error that emerged was that of pilots fixating on the GPS display to the exclusion of aircraft control, sometimes showing dangerously large deviations in altitude and course. Fixation errors are of critical importance in the current flight training environment as modern cockpits utilize more sophisticated displays and avionics.

Based on the findings from this study, the following recommendations for future research are presented:

- The same experiments could be conducted with students at a different flight school to enhance external validity.
- The experiment could be conducted with a longer time period between pretest and posttest flights and controlled for extraneous variables to test for long term training effects.
- A time-series design could be used to determine longer term training effects.
- A regression-discontinuity design might describe training effects for pilots who initially performed lower and those that performed higher.
- A qualitative study using responses from participants could comment on the extent of learning that occurred.
- A larger sample size could be used to enhance external validity.
- A study could be designed to test pilots in cockpits with various levels of complexity, for example using one of the many new general aviation flat panel PFD/MFD or virtual 3D displays installed in many new aircraft.
- A cost to benefits analysis could be conducted to determine if task prioritization training should be incorporated as a component of a training course outline.

References

Bishara, S. and K. Funk (2002). Training pilots to prioritize tasks, *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, Baltimore Maryland, September 30 – October 4, 2002, 96-100

Chappell, S. L. (1998). Disciplined attention, the next best thing to a crystal ball. In T. Kern, *Flight discipline*. New York: McGraw-Hill.

Chou, C.D., D. Madhavan, and K.H. Funk (1996). Studies of cockpit task management errors, *International Journal of Aviation Psychology*, Vol. 6, No. 4, 307-320.

Federal Aviation Administration (1999). *Aviation instructor's handbook* (FAA-H-8083-9). Washington D.C.: U.S. Government Printing Office.

Funk, K., B. Lyall, J. Wilson, R. Vint, M. Niemczyk, C. Suroteguh, and G. Owen (1999). Flight deck automation issues, *International Journal of Aviation Psychology*, Vol. 9, No. 2, 109-123.

Funk, K. and R. Braune (1999). *The AgendaManager: A knowledge-based system to facilitate the management of flight deck activities*, SAE 1999-01-5536, SAE International, Warrendale, PA, presented at 1999 World Aviation Congress, San Francisco, CA, October 19-21, 1999.

Jeppesen (2002). *Flight instructor manual*. Englewood, CO: Author.

Jeppesen (2003a). *Instrument commercial manual*. Englewood, CO: Author.

Jeppesen (2003b). *Private pilot manual*. Englewood, CO: Author.

Kern, T. (1998). *Flight discipline*. New York: McGraw-Hill.

Kershner, W. K. (1998). *The instrument flight manual* (5th ed.). Ames, IA: Iowa State University Press.

Machado, R. (2001). *Rod Machado's Private Pilot Handbook.* San Clemente, CA: The Aviation Speaker's Bureau.

Machado, R. (2003) *Rod Machado's instrument pilot's survival manual*. San Clemente, CA: The Aviation Speaker's Bureau.

Thom, T. (1991). *The pilot's manual: volume 2 flight operations.* New York: Center for Aviation Theory.

Wilson, J. R. (1998). *The effect of automation on the frequency oft task prioritization errors on commercial aircraft flight decks: An ASRS incident report study*. Unpublished master's thesis, Oregon State University, Corvallis.

TASK ANALYSIS FOR SAFETY ASSESSMENT IN EN-ROUTE AIR TRAFFIC CONTROL

Satoru Inoue
Graduate School of Frontier Sciences, The University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN

Hisae Aoyama
Traffic Management Systems Division, Electronic Navigation Research Institute
7-42-23, Jindaijihigashi-machi, Chofu, Tokyo 182-0012, JAPAN

Kota Kageyama
Traffic Management Systems Division, Electronic Navigation Research Institute
7-42-23, Jindaijihigashi-machi, Chofu, Tokyo 182-0012, JAPAN

Kazuo Furuta
Graduate School of Engineering, The University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, JAPAN

The tasks involved in Air Traffic Control (ATC) make heavy demands on the information processing capacities of air traffic controllers. In particular, human factors problems that lead to both major and minor incidents are considered to be a serious problem for ATC in Air Traffic safety. Therefore, the need to analyze error mechanism, which occurs due to complex factors, and the need for developing systems that can deal with these errors are increasing. We examined the functional problems in an ATC system from the human factors aspects, and concluded that solution of this problem needs some kinds of measures. This research focuses on analysis of the air traffic controller's tasks for en-route ATC and modeling controller's cognitive process.

Introduction

Recently the workload of Air Traffic Control (ATC) has become heavier due to the increase in air traffic demands. Especially human errors that lead to both major and minor incidents are considered to be a serious problem for air traffic safety management. Human factors problems in ATC can be observed or tackled from various aspects. However, little has been known about the causal factors leading to human errors in the current ATC systems.

Thus, we need to understand details of basic functions of air traffic controller's tasks in the systems, in order to design more reliable interfaces or training programs for the controllers. This research focuses on task analysis of air traffic controllers in actual en-route ATC in an experimental approach. We first discuss the idea behind the experiment relying on principles of ethnomethodology, and then show some findings obtained from the experiment.

Approach

ATC is a very complex process that depends to a large degree on human capabilities. The design of advanced and efficient ATC systems for the future requires understanding of the nature of interactions between the controller and the basic available sources of information such as the radar display console, paper flight progress strips, aircraft pilots, and other controllers.

In order to design the system that can assure system safety, enhance usability, and support human reliability in the future, it is critical factor for an developer's engineer to consider the feature in the control system operation and the intention of the controller.

An effective method to understand user's requirements is to analyze user tasks based on actual field data. This research aims to make a model of cognitive process of an air traffic controller through task analysis, to find the problem from human factors perspective for improving design of future air control systems.

Suchman (1987) pointed out the need of an ethnographic approach on the site of work when it is the problem what knowledge and experience people use in a cooperative work. Ethnomethodology is a method of sociology to find out some implicit orders, rules, or norms behind human activities through observation in the actual work environment (Ando, 2003). Both the research of works in the cockpit of aircraft by Hutchins (1994), and the research of works in the London underground line control center by Heath and Luff (1992) are based on this idea of ethnomethodology.

We believe ethnomethodology is one of the effective methods for analysis of ATC tasks, because when we examine human factors problem, it is important to understand the actual work environment. In this research, an experimental task analysis was done by an ethnographic approach.

Experimental Setting

To analyze how air traffic controllers work, we built an experiment system for collecting data through simulator experiment. In the experiment, we recorded motions, sounds, and simulator logs as basic data for the analysis. From these basic data, we reconstructed controller's actions and protocol logs, and analyzed controller's tasks in each situation. The system has functions to record multiple types of time-series data such as video, audio and simulator logs.

Figure.1 shows the setting for video and audio recording in the simulator room. Fixed cameras record actions of the controllers, displays of instruments, the chart table, and projected situations from four angles. Three capacitor microphones can record all communications between the controllers in the room and communication for pilots.

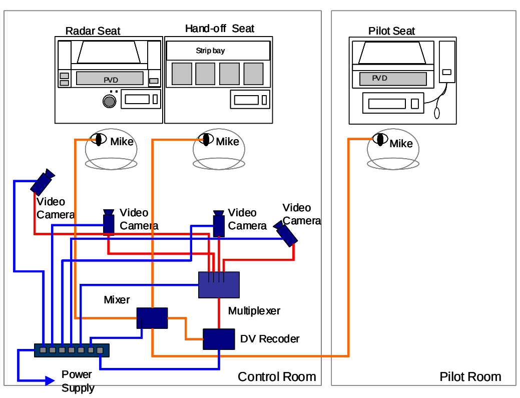


Figure 1. *Setting of experiment*

- Video (VTR) data - The VTR records air traffic controller's behavior like instruction, coordination, etc., in the control room. Four cameras record the situation of the entire control room from four directions including the radar screen, the flight-data-strip bay, the seat of radar controller, and that of coordination controller. Moreover, this system combines videos of the all cameras and audios from the radar controller, the coordination controller, and the pilot, takes synchronization, and then records them in a batch.
- Flight-data-strip - As for flight-data-strips, marks, notes, and so on are written down by the air traffic controller during controlling work. We understand clearance and instruction for individual airplane pilots and the content of coordination to other sectors from the records on these papers.
- Simulator track log - All the route patterns and parameters of each aircraft during experimental runs are recorded in the simulator as a log. These records together with records of video and flight-data-strips are used to understand behavior of the air traffic controller, the control situation at each moment, and consequence of controller's instruction.

Method of Task Analysis

The radar controller and the coordination controller, who takes charge of en-route ATC, frequently monitor the display of the radar control interface and data of flight-data-strips, and carry out controlling tasks while exchanging information. For instance, when the radar controller projects existence of a related aircraft from the radar monitor, a series of works of the radar controller is directed to the pilot by communication of an appropriate instruction to the aircraft to avoid conflict. The controllers then input the content of instructions to the RDP (Radar Data Processing) system, and mark the flight-data-strip.

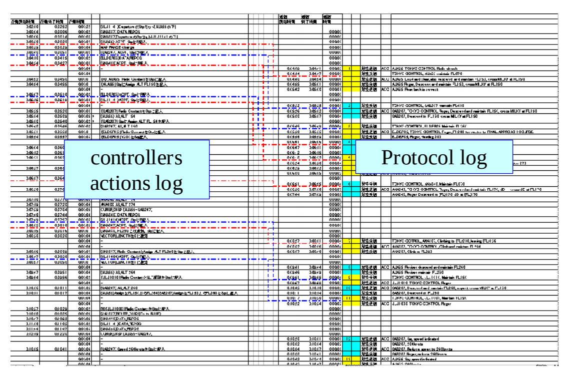


Figure 2. *Data logs in a time line*

A sequence of controller’s tasks are described into time line data that consist of action log and protocol log as shown in Figure.2 based on each data of videos, flight-data-strips, and simulator logs recorded with the experiment system. The situation is then segmented following the content of radar controller's communication mainly based on the time line data of action and protocol. The context of each segmented situation is analyzed based on the action and protocol data as well as the explanation of situation made by a supervisor (Figure.3). In addition, Goal-Means Task Analysis (GMTA) was performed on these situations (Hollnagel, 1993).

3:07:18	3:07:22	5		TOKYO CONTROL ENJOY01 270
3:07:22	3:07:26	5	ACC	ENJOY01 TOKYO CONTROL ROGER SAY
3:07:26	3:07:29	4		ENJOY01 HEADING 143
3:07:29	3:07:36	8	ACC	ENJOY01 TURN LEFT HEADING 130 DUE TO TRAFFIC EXPECT VECTOR TO FINAL
3:07:37	3:07:40	4		ROGER HEADING 130

<Explanation>
ENJOY01 was approaching to RJAH. The controller did not want it to go through GOC for avoiding conflict with AUA51. The controller used a vector as the method of control and directed ENJOY01 to Heading 130 here.

Figure 3. *Example of situation segments*

Result of Experiment

Conditions

The Kanto North sector shown in Figure.4 was used as the target of simulation experiment, and the subjects are professional controllers qualified for this sector.

We monitored behavior of a team of a radar and a coordinator controller working on a simulation scenario of about 60 minutes where they performed multiple tasks of handling many planes at a time. The controllers controlled 75 aircrafts in an hour. Table.1 shows the detail of the traffic handled in the scenario. The data of the first eight lines in the table correspond to the aircrafts that require climbing or descending instructions from the controllers in compliance with the control regulation in the sector. The controllers do not need to intervene these aircrafts as long as enough separation is kept. The number of over flights is three in total out of 75 aircrafts. It is the feature of this sector therefore that the major traffic is a flow of aircrafts climbing from or descending to airports. The amount of traffic assumed in the scenario is relatively heavy.

Table 1. *Detail of experimental scenarios*

	Bound	
Airport	DEP	IN
RJTT (TIA)	14	22
RJAA (NRT)	6	9
RJSF	3	2
RJTY RAPCON	5	4
RJTK	1	0
RJAH	4	1
RJTU	1	1
RJSS, RJSC	2	3
OVER FLIGHT		3

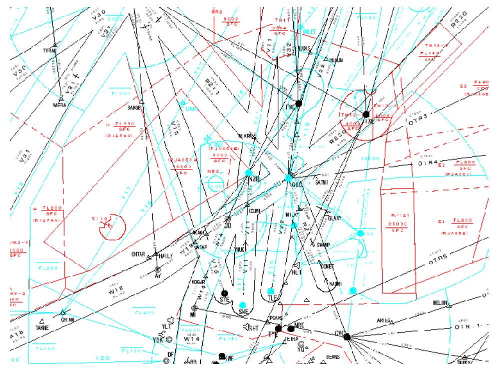


Figure 4. *Map of Kanto North sector*

The Data of Result

We have already finished analysis of data for three teams at present. Table.2 shows the number of communications and situation segments from the communications.

Table 2. *Number of communications and segments*

	Subject team		
	A	B	C
Number of communications	598	567	582
Number of segments	256	234	259

Some differences exist in the number of communications depending on the content or the way of communication. Instructions were issued for many purposes: initial contacting, clearance, spacing, radio frequency transfer (hand-off), etc.

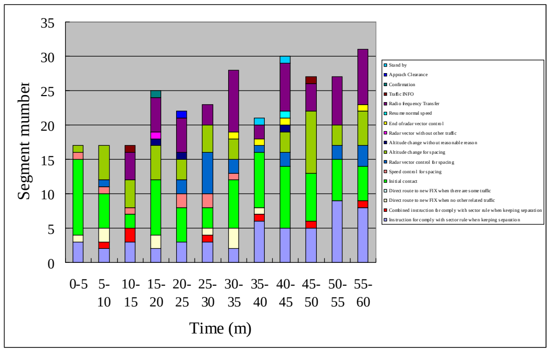


Figure 5. *Content of communication contents*

Figure.5 shows the number of situation segments totaled by every five minutes and classified by the content of communication. The peak of traffic comes in this experiment at 25 minutes and 40 minutes. We can see that many instructions for the spacing are concentrated on during that time. It can be understood that the radar controller put out a variety of control instructions along the situation for spacing. However, we do not understand the radar controller's cognition and decision making process from such a statistical method.

Case Analysis

It is difficult to understand the decision making process of the controllers how to decide particular instruction in particular situation. We analyzed an individual process of decision making that resulted in a single segment of communication. Since too many segments exist as shown in Table 2 to show every result of such analysis, one example will be given here that well reflects the geographic features and the regulation rules of this sector.

It is characteristic that a lot of aircrafts come into this sector from more than two sides of northern sectors to land at the Tokyo International Airport (TIA). The controllers should guide these aircrafts down to an altitude of 13,000ft by TLE, which is the point to transfer descending aircrafts to TIA RAPCON (radar approach control) and to handoff to the next sector, while keeping separation above 10 mile in the trail. The way and the content of instruction to aircrafts from more than two directions are important for the control tasks in this case. For instance, let us think about the relating situation of four aircrafts shown in Figure.6 from the experiment. In this situation four planes are coming from three districts in the north aiming at TIA. The controllers have to line up these four aircrafts at 13,000ft and keep separation in 10nm each aiming at TLE.

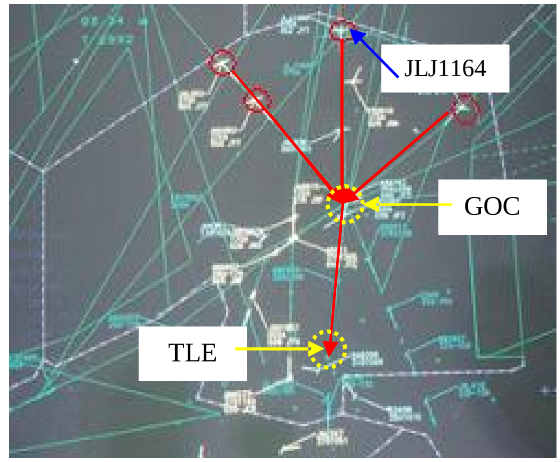


Figure 6. *Example situation for case analysis*

The radar controller directed the instruction "JLJ1164 FLY HEADING 170 FOR SPACING" at Figure.6. Figure.7 is a description of the situation to JLJ1164 at the situation that is shown in Figure.6. At this time, the radar controller considers JLJ1164 that is No.3 in-bound to TLE, and directs an instruction for spacing as the way of radar vector control.

Target:< To keep the safety interval of JLJ1164>
Task: To keep the safety interval.
Precondition: To understand the status of the aircraft.
Target:< To understand the status of the aircraft >
Task: Radar monitoring or confirmation by communication
Precondition: Clearance of related traffic.
Target:< Related traffic is cleared. >
Task: To examine the method of spacing.
Precondition: To examine the control method to the destination.
Target:< To examine the control method to the destination >
Task: Route retrieval.
Precondition: To examine the method according to the control rules.

Figure 7. *GMTA of the case*

The result of GMTA on one arbitrary segment related to the case situation is shown in Figure.7 as a schema. We can thereby understand the process of radar controller's decision making in one situation. GMTA defines descriptions of the task step and the feature of the tasks of the action in specific situations. We can observe the common performance modes of the controller by GMTA. The common performance modes are a convenient way of describing the impact of the context on the control of actions. The common performance modes can be determined from the outcome of the GMTA (Hollnagel, 1993).

We recognized the state of the controller's cognitive and decision making process as a model (ex: depth of the situation comprehension, accuracy of the projection) in a situation of the individual segment in this experiment. Especially, It can be expressed the relation between control mode of the controller's performance and cognitive process of the controller in this analysis. This result of analysis shows the level of the cognitive process model in individual tasks of the controller in detail. This schema, however, is for the situation of a single segment.

A comprehensive model is necessary to express a series of cognitive process until handoff is done. Situation analysis by observation and interview of the controllers can be repeated to reveal a series of cognitive process.

The strategy for each situation of an individual radar controller does not differ greatly, because conditions are limited from regional characteristics and the rules of the sector. Concrete methods, however, of applying the strategy had some individual variations. When keeping separation, for instance, one controller used speed adjustment several times without removing aircrafts from the route, but another controller used vector instruction rather than speed adjustment from the beginning.

Cognitive Model

Features of Tasks

This chapter describes construction of a cognitive model of an air traffic controller from the observation and analysis of the experimental records. Kawano (2001) mentioned that there are some specific features in work of ATC. Especially the basis of the work is prediction and instruction to secure a safe situation in the future.

As for the radar controller in en-route control tasks, he/she predicts from five to ten minutes ahead. Meanwhile the coordination controller elaborates instruction to keep safe separation in the previous state from the information available at present.

A lot of interruptions will happen when the controllers have to handle more than two aircrafts at the same time: call from another aircraft than that of current interest, request of hand-off from another sector, and so on.

The coordination controller has similar tasks with interruptions to keep coordination with next sectors. In addition, the controllers have to control all IFR aircrafts in their own sector. Since en-route ATC work have to deal with a variety of states and conditions of the sector, it differs greatly from well formalized tasks like assembly line operation.

Control Mode of ATC Controllers

The model presented here represents the routine task of decision making and performance of the radar air traffic controller as a flow chart. When issuing a conflict avoidance command to keep separation of airplanes, the priority is determined from the relative distance and velocity of airplanes, an appropriate avoidance method is chosen from the flight situation, and then the instruction is given to the airplane. The control modes of a controller in the above process can be defined based on the Contextual Control Model (COCOM) of Hollnagel (1993) shown in Figure.8. COCOM consists of four control modes of human performance.

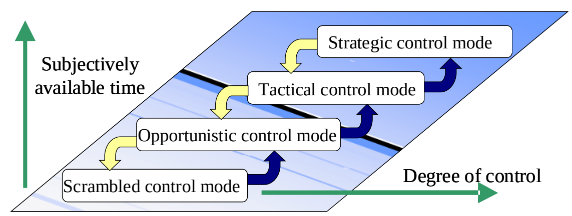


Figure 8. *Control modes of COCOM*

Usually the air traffic controller is working within a range from the strategic to the tactical control mode. It is well known that air traffic controllers are likely to err in the opportunistic mode, because they will take an action based on its face validity of situation without profound awareness. Talking about the control mode of each air traffic controller, the radar controller is almost in the tactical mode, because the time margin for his/her decision making is relatively restricted.

Cognitive Process of Radar Controller

In this research, we examined the basic cognitive process of the radar controller in a state of a single task. Controller's tasks are restricted by geographical features of the sector, the air route characteristics, the control rules, etc. In addition, the controllers are highly trained to handle the tasks efficiently and safely in a very restricted time interval.

The decision making process of an air traffic controller is defined as the model shown in Figure.9 from the observation of the experiment and the analysis of interview to the subjects. This basic model follows Endsley's model of Naturalistic Decision Making (Endsley, 1997).

This process will arrive at decision through search of the target by perception (perception), understanding of the sector situation (comprehension), prediction of the future state of aircrafts (projection), and execution of action.

It is highly depending on time margin available for each process whether the process of decision making is strategic, tactical, or opportunistic. Instruction becomes strategic if there is a lot of time margin in the all processes. We observed that the content of judgment could sometimes become unrelated with the time margin when short cut of the process happens by heuristic situation assessment in each process.

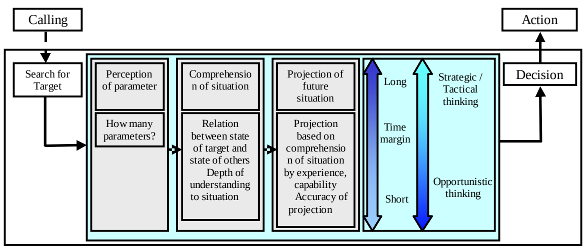


Figure 9. *Cognitive model of radar controller*

The radar controller executes such a cognitive process in a very short time. It seems that the experience of the controller has an important effect on his/her situation projection in this state. The air traffic controller has a model of situation assessment originated from his/her experience. We obtained an expectation that the controllers made a decision in this experiment by pattern matching with the data base of the model.

Conclusion

In this research, we proposed a technique for analyzing tasks of ATC by a method of ethnomethodology as an approach to study problems of human factors in an ATC system. We will continue the data analysis to understand detailed features of cognitive process of a controller team. We are going thereby to construct a model of team cognitive process and then a database for both the radar and the coordination controller.

References

Ando, H., Yamato, H., Koyama, T., Kakuta, R., Nishida, H. & Enomoto, M. (2003). Discourse Analysis of Crew. In proceeding of International Conference on Marine Simulation and Ship Maneuverability 2003,

Endsley, M. (1997) . The Role of Situation Awareness in Naturalistic Decision Making. *In Naturalistic Decision Making, 269-283.* Lawrence Erlbaum Association, Inc.

Heath, C. and Luff, P. (1992). Crisis Management and Multimedia Technology in London Underground Line Control Rooms. *Journal of Computer Supported Cooperative Work, Vol.1, No.1,24-48*

Hollnagel, E. (1993). *Human reliability analysis - Context and control.* Academic Press Ltd.

Hutchins, E. (1994). *Cognition in the Wild*. MIT Press.

Kawano, R. (2001). Facts of Human Error in Air Traffic Control. Human Interface (JAPAN),*Vol.3 No.4 , 221-228.*

Suchman, L. (1987). *Plans and Situated Actions.* Cambridge University Press.

PILOT PERSPECTIVES ON AVOIDING CFIT

Robert M. Jenney
Aviation Safety Connection, Inc.
Reading, MA USA

According to Flight Safety Foundation, Controlled Flight Into Terrain (CFIT) accidents are a leading cause of commercial aviation fatalities. CFIT is not selective; it is prevalent in corporate aviation and, all too alarmingly, in other segments of general aviation. This paper is a pilot-to-pilot conversation specifically about avoiding CFIT and, in the process, about airmanship. All are invited to listen in.

The Conversation

I know the CFIT statistics and the devastating consequence of this type of accident but, as a pilot interested in an aviation career, what can I do to avoid CFIT?

That's a difficult question to answer simply. The body of research is large, much is of the information is based on actual events and accidents, and good ideas result. But, the end product doesn't always produce specific guidance—that task is often left for the training departments, instructors and individual pilots to accomplish.

Let me draw upon the experiences and beliefs of other pilots, inject my own biases, and examine some of the thought processes that have proven to be useful. However, in the end you will need to establish your own guiding principles and personal standards on this issue and others that will define you as an airman.

How does avoiding CFIT fit into this concept of airmanship?

First, let's set the parameters. What is CFIT? Flight Safety Foundation defines a CFIT accident as, "one in which an otherwise serviceable aircraft, under the control of the flight crew, is flown unintentionally into terrain, obstacles or water, usually with no prior awareness on the part of the crew of the impending collision." In short, CFIT is loss of near-ground position awareness; the airplane is controllable; the results are fatal.

Position awareness is not a momentary, static geographic plan view. It's a three-dimensional understanding of where you've been, where you are and where you're going to be. But, we can't simply narrow our focus to positional status to grasp and deal with this CFIT problem. There are many factors involved, and we need to broaden our outlook to the more encompassing concept of situation awareness (SA), of which near-ground awareness is a specific case.

And the complexity of this broader view demands a high degree of professionalism—airmanship—from pilots. Airmanship requires confident piloting skills, sound judgment and a strong sense of personal accomplishment and well-being. Even with the finest technical flying skills, poor judgment leads to faulty decisions and, very likely, to loss of SA.

I'd been thinking of awareness in terms of location and position. What else is involved?

The varied aspects of flying are widespread and interrelated. Physically they include the aircraft, its flight, operating, navigation and communication systems; the aviation system, its traffic control, its navigation/communications network and navaids; its airports, runways and taxiways. Environmentally they include weather and its uncertainties; severe weather with its icing, storm cells and microbursts conditions; restrictions to flight from low ceilings and limited visibility; turbulence, winds, night conditions, sun position, etc. In the cockpit our concerns are all the flight parameters (altitude, airspeed, angle of attack, G-forces, flight progress, fuel status, etc.). Personally we deal with aeromedical issues such as spatial disorientation and illusions, vertigo, hypoxia, stress and fatigue, and ergonomic factors resulting from equipment design, control and switch locations, etc. Plus, personal interactions are many and varied.

These factors influence how well we fly and, separately and in combination, determine the situations we need to be aware of.

Situations change, and maintaining awareness in flight is a dynamic process. In fact, it is an integral part of what you know as "aeronautical decision making (ADM)."

OK, so it's a bit complex. But I'd really like to focus on my original question. Can't we get back to CFIT?

Actually, no. Managing risk requires understanding human fallibility and loss of SA, and understanding requires effort and patience. We all want to climb in the cockpit and demonstrate our piloting ability, but the truth is we may not be as good as we think. There are a number of good and experienced pilots who have been victims of CFIT—it only takes one mistake, and every pilot I know admits to mistakes.

Let's continue. ADM is the process of judgment and decision making. We exercise judgment based on our experiences, our biases and pressures we place on ourselves. Any change that's perceived requires a decision. That decision may be to take any number of different actions, or may be not to act at all. The quality of your judgment—the option you choose—is measured by the resulting decision you made. But that's not the end. Decisions require action, actions induce change, and change necessitates judgment. It's a continuing feed-forward, feedback process.

I'll stay with you and try to be patient. Judgment is selecting options, decision making requires action. Where does SA fit in?
Remember I said that in terms of SA, pilot's think of where they've been, where they are and where there're going to be? Looking ahead requires anticipation and some degree of expectation.

At this point let me say that I've talked with a number of pilots and conducted informal surveys on issues related to SA, and feel I can speak with some confidence. You should make an effort to do the same because there's so much that can be learned from peer interaction.

Another pointer. The first thing I read in a magazine is letters to the editor—here you get the real-world opinions. Recently, Professional Pilot conducted a pilot survey on SA and published the responses.[1] If you carefully read comments like these you'll get a good sense of how other pilots think. You won't always agree but, again, you'll need to form your own opinions over time.

Back to your question on where SA fits in. We need to get beyond the words and put them in the context of how you might think and act.

On looking ahead. Anticipating. We prepare for each flight. We flight plan and consider any number of factors including weather, winds, fuel consumption, alternates, approaches, airport characteristics—our situation list. We brief ourselves and, perhaps, others. And, we brief in explicit terms, not abstract thoughts. We develop expectations of what lies ahead.

In flight we alter our expectations to actual events. How we prepare for and adapt to change varies with each individual, but I'd like to suggest that a structured thinking process that can be relied upon is desirable, particularly under conditions of high workload. Stress is relieved when we're confident in our ability to anticipate problems and deal with them when they arise. Let's look at how some pilots think as they manage their cockpits. They play "what if?".

Only it's not a game. It's a serious management method that looks at an uncertain future and postulates alternative versions of that future. It works in business and works in the cockpit.

"What if?" Scenario building. Let's see how "what if?" could work in developing in-flight strategies.

1. Select an issue that you may be confronted with. A few possibilities: fuel (reserves or consumption); weather (enroute, at destination, at alternate); winds aloft; clearance; equipment.
2. Think of two or three possible outcomes ("what if" scenarios).
3. Think through the implications and work out a game plan (strategy) for each scenario ("what if—then…").
4. Establish a time or place (fix) to review your issue.
5. Revisit and revise, as necessary, your scenarios and game plans.

The "what if—then…" process is simple and straightforward, more difficult to describe than utilize. Through these steps you have put certain issues in perspective. You've gained greater awareness of the issues facing your flight. You haven't played your hand, but this disciplined form of thinking better prepares to make choices when you need to. You won't have "covered all the bases," but you will be able to deal more fully with new issues that pop-up since you've got a few others covered.

These scenarios can be developed by the single pilot, of course. They will tend to be more thorough in a cooperative two-pilot team cockpit. That's at least an introduction to options and SA. Are you with me?

Yes. I can see that working out in-flight strategies promotes increased awareness. It also lays out a few options that might be considered at an appropriate time. What about decision making and SA?
Other pilots will respect you for your flying ability, but they will judge you more critically by the decisions you make. We'll discuss and emphasize a few of the finer points, but I'd like to emphasize what I consider to be key elements. Know your airplane. Know your mission, visualize it and prepare for it. Then you'll be ready for in-flight changes when they occur. There's seldom need to rush, so approach your tasks systematically. Define and set your goals in

precise terms. Understand that if your expectations are not clear and your targets vague, then the quality of your resulting decisions and actions will suffer.

On maintaining control. Making corrections. We want to keep control of our activities and decisions as smoothly as we control our aircraft. The familiar feedback loop is at work—input, output, feedback, correction. It's exactly what you do or your autopilot does for you. Your cockpit management input is what you want to achieve, your objective, goal or target. Your output is what you're actually accomplishing, your result. The difference between what you want and what you get is the information feedback you need to act to get back on target. Clearly, the smaller the deviation the smoother the correction.

Before we move on, I'd like to emphasize two important points. One, be committed to your objective. If you waver in this commitment, then you've given up a measure of control. Your autopilot is tireless in its commitment to your input. You need to "stay the course" on your cockpit management activities as well. Two, set boundaries or tolerances with respect to your objective. These may be rules set by your organization or guideposts that you've established for yourself. How much variation are you willing to accept? For example, we know that unstabilized approaches lead to CFIT accidents. Define clearly and set in your own mind what a stable approach is. Perfect entries are not always possible, but know at what point you must be stable (within your predefined tolerances) or break it off. Stay in charge—control your destiny.

On ADM. Judgment and decision making. Flight conditions change, and change requires judgment be exercised. Almost all judgment is intuitive, based on experience, training, personal beliefs, professional standards, and the ability to determine right from wrong. If it weren't intuitive we'd be bogged down and unable to function.

The balance of judgment is cognitive. We make a conscious choice of what to do. It follows, then, that we want to elevate the more crucial situations to this more thoughtful level and base our choices on conscious evaluation. To do this, we look ahead, determine viable options, anticipate outcomes and select courses of action. We're interested in exercising sound judgment, not necessarily the very best, since it's impossible to evaluate and rate all possible options.

Decision making is judgment's product. The decision's objective (target) is defined, progress towards that objective is monitored (feedback), and corrective action is taken when necessary to keep the target in sight. At any one time there might be multiple decisions in play, so feedback is apt to be intermittent rather than continuous. The better pilots are able to keep them all in perspective, checking the flow of information for each sufficiently to maintain a state of management control.

It is this information flow—feedback of progress towards the goal—that establishes and maintains awareness. That's why I say that SA is an integral part of a quality decision process, not a separate, stand alone characteristic. Good decision making enhances awareness.

We've pretty well summarized the general topic of situation awareness. Now let's look at the more specific nature of near-ground awareness.

Good. I've been waiting for this!
Awareness of your surroundings is all the more crucial at lower altitudes. This is a good time to cover one other aspect of SA, self awareness. Add it to your situation awareness list—it's probably your list's most important item.

Self awareness. Self knowledge. We've all been taught aviation's cardinal rule: "Keep Flying the Airplane." We accept its premise. We understand its logic. Yet, it's surprising how often this simple rule is violated and an accident results. We also know we should act immediately and climb above minimum safe altitude (MSA) when near-ground position awareness is lost. But, as you know from your statistics, too often pilots don't take this action.

Knowledge of these fundamental behavioral rules is not enough. It's not an intellectual exercise. Are you committed to acting? Will you really do what you know is right at that crucial moment in time?

Dig deep! Imbed rules such as these in your personal belief system. Commit yourself to act. "Know Thyself" as the ancient axiom dictates. Know that you will act appropriately and decisively when the time comes. Acquitting the pilot in command's responsibility and authority and knowing "the consequences of your actions"[2] requires this type of individual dedication.

Thanks for the advice. I hadn't thought of myself as a pilot in quite those terms.
You're welcome. It's easy to fall into the trap of thinking, "Let's see what develops—I can handle it." On the subject of CFIT, there's another truth: "In the battle with Mother Earth, the lady always wins."

Let's move on and cover some CFIT topics. You know the other contraction, CFTT (for controlled flight towards terrain). No accident results, but an accident might have occurred if evasive action had not been taken. These are real learning experiences that unquestionably have changed the subsequent behaviors of those involved. One pilot told me:
"I realized it just in time. It was scary, but there's been instant improvement of SA from then on."

They can be learning experiences for you as well. A search of the ASRS database using "CFTT or CFIT" will produce well over 100 incidents. If you selectively read a few, get a real sense of the circumstances and mentally place yourself in the cockpit, you'll be able to learn a great deal and benefit from the mistakes of others.

And talk to other pilots. As I said, there's a huge database of aviation knowledge in the minds of others who do what you do. In my informal survey I found that, on a pilot-to-pilot basis, pilots are willing to share their experiences openly and honestly. We don't have this type of conversation often enough.

I agree. I've learned a great deal from my colleagues, and it's a great feeling to know I can call on more experienced pilots for advice. So many are willing to be informal mentors.
That's one rewarding aspect of our industry. As I said earlier, pilots are usually forthright in acknowledging mistakes: "Be aware that everyone makes mistakes and cannot operate at their best on all occasions."
Now, to a few specific CFIT/CFTT topics:

CFIT themes. Crew behavior patterns. One pilot who is concerned with the CFIT issue sent me his observations, as follows:[3] "From my experiences of investigating CFIT accidents I have seen the following common themes involving situation awareness and crew monitoring.

A. The crews saw something—the ground or non-aeronautical lights were misinterpreted. In most of these accidents, the crews were not adhering to the definitions of MDA/IMC, or not aware of the precise visual requirements for a land decision.
B. Both crew members were comfortable with their navigation position and/or their actions, they both made the same assessment and/or the same mistake, thus the cross monitoring function failed.
C. The circumstances did not enable any monitoring; the monitoring pilot remained head down and could not see what the captain was describing (ground contact). The flying pilot had made an error, but the monitoring pilot did not/could not/would not detect it.
D. There are covert peer pressures due to the expectations of the industry. I.e., a go-around carries a professional stigma—ATC sees it, other pilots comment, management wants to know why. In these accidents the crew formed an opinion of the airfield/flight conditions and briefed for that plan; they were unable/unwilling to change their plan as the real conditions unfolded—situation awareness/decision-making."

To be effective, CFIT prevention needs to take into account these and other identifiable themes and derive appropriate countermeasures. It's advisable to keep in mind that many CFIT accidents occur near airports over relatively flat ground, not necessarily rugged terrain.

Another good introduction to CFIT is the Flight Safety Foundation's video, complete with accident recreations.[4] It reports that fatalities are greatest in the transport category with air taxis providing the highest rate, and that frequent accidents occur in the approach phase with multiple step-down approaches being prominent.

When you view the video it you will see the effects of scud running (by two very experienced pilots), an accident due to black hole illusion on a straight-in approach, and the result of confusing ATC clearances with communications conducted by parties who had different primary languages. The concluding message: acknowledge vulnerability and be vigilant.

You seem to have given me homework assignments!
Strictly voluntary, but passive classroom attendance isn't enough to understanding the problem and developing your own behavioral guidelines. Now, let's be more specific.

Personal standards. Integrity. We started this conversation during the self awareness discussion. Here are a few of the specific commitments pilots make to themselves when regard to terrain avoidance:
"You really have to stick to your standards and not cave in to pressure from others in the cabin."
"I never do circling approaches at night in low visibility if there are any obstructions in the area."
"Go-around or climb if you're ever in doubt."[5]
"A landing is an approach without a go-around, mentally preparing the GA as the escape maneuver."
"A go-around is present in my mind throughout the approach. In fact, go-around is my aim until the situation is suitable for landing."

The concept that landing is a go-around interrupted by a decision to land is a valuable insight that appears to be gaining acceptance.

Preparation. Flight planning. The need for mental preparation is obvious, but a quick review doesn't hurt. Plus, there are a few good pointers in pilot comments and thoughts to keep in mind:
"Even when solo I give myself a detailed briefing, particularly on airport characteristics and obstacles."
"Gather information, as much as you can, about the flight to improve situation awareness."
"I study the VFR sectional, even though most of my flying is on IFR flight plans."
"I look for certain clues to the presence of obstacles—displaced thresholds, circling minima that vary with the category of the aircraft, circling restrictions, departure procedures."
"Non-precision approaches need special briefing attention."
"Being an east coast pilot, I think any airport above sea level deserves my full attention."

Pilots understand the need to be aware of hazard potential. However, many don't think of CFIT as distinct and separate issue since, with proper preparation, terrain and obstacles are part of the normal flight environment:
"My concept is that CFIT/terrain awareness must be embedded in everyday activities and is not a 'special' or bolt-on activity."
"Most of my routes are into areas of high ground and prone to heavy rain and poor visibility, so terrain separation is an every day exercise."

Options. Judgment's choices. Here are a few additional pilot thoughts that complement our earlier scenario development topic:
"Think "what if" and apply everything you've learned."[6]
"I always ask myself what I would do if a situation happened right now."[7]
"My basic principle is always have a solid gold plan B, maybe a plan C as well, and not to let risks compound with each other."
"Review all aspects of descent and approach during cruise and be sure to discuss all options."[8]
"Good SA involves being both physically and mentally aware … and what options are available should something go awry."[9]

Well, I seem to be getting good points to think about.
Yes, I think you are from your fellow pilots. Next, you'll get their view of a really critical issue, stable approaches, and setting specific targets during the approach to maintain cockpit management control.

Stabilized approach. Crucial to terrain avoidance. Inviolable rules and clear thinking are the hallmarks for maintaining control during approaches.

We won't get into great detail here, but pilots do establish defined tolerances for approach entries, the point the aircraft must be properly configured, the point at which they are stable within their boundaries (speed, rate of descent, +/- GS/LOC dots, etc.). A few related thoughts, but we're barely touching this important subject:
"A stabilized approach is my safety net."
"On a typical ILS approach, I get it fully configured and stabilized by 800 AGL or earlier. On non-precision approaches I try to be fully stabilized beginning at the FAF."
"I adhere to the stabilized approach philosophy and approach gate concept."

When flying alone, many pilots brief themselves and make altitude callouts through the intercom. It's not loneliness—it comes from the knowledge that speaking out loud makes each thought more concrete and specific, less abstract.
"During an approach I try to verbalize what is happening, even to myself if I am alone in the plane."

And, as with "what if?", asking and answering questions is a proven method of defining the situation more clearly than observing events taking place. Here is how one general aviation pilot maintains a stable approach and position awareness:
"I think about each segment of the approach using three questions: (i) 'how low?'; (ii) 'how far?'; (iii) 'what's next?'." By answering these questions he's establishing specific targets that he'll be able to measure his performance by.

We can all learn from this pilot's thoughtful and thorough technique. Few pilots I know have developed as comprehensive method for asking and answering questions that give confidence they're in control of their flight.

I do see the value of Q & A for position awareness as well as the earlier development of scenarios. I'll give techniques like these a great deal more thought.
Good. Let's briefly mention a few more topics related to CFIT/CFTT and perhaps expand on them later.

Sterile cockpit. Minimizing distracts. It's said that regulations are "written with blood" as an after-the-fact response to dramatic events. That's probably how this concept became a requirement for Part 121 cockpits. The idea is to aid SA by minimizing distractions and interruptions during crucial phases of flight and, although not required, others have adopted versions of the sterile cockpit concept.

Here's a summary of the regulation: No flight crewmember may perform any duties during a critical phase of flight except those required for the safe operation of the aircraft. Critical phases of flight include all flight operations conducted below 10,000 feet, except cruise flight.[10]

And, here are ways that sterile cockpit is interpreted:
Transport category: Sterile cockpit below 10,000 feet (per regulation).
Corporate training program: "A sterile cockpit will be maintained during dynamic (non-cruise) flight."[11]
GA pilot: "I observe sterile cockpit rules within 40 miles of the destination."
Helicopter pilot: "Keeping good SA involves disregarding distractions such as passenger chatter and staying connected with what's going on inside the cockpit."[12]

Sterile cockpit is an important rule for all near-ground operations—it's certainly advisable for general aviation pilots to conduct this briefing with their passengers.

Minimum Safe Altitude (MSA). A primary SA threshold. It's the altitude that's established to provide at least 1,000 feet of clearance above the highest obstacle in a near-airport sector. Pilots develop rules and have definite thoughts about how to use this information, a few of which follow:
"The MSA is the tip of the iceberg on being aware of what hazards lie around the airport."
"Do not fly below MSA unless under radar control or established on a charted approach, or visual–VMC."
"Situational awareness is knowing … your MSAs and having all cockpit instrumentation set up for what's going to happen over the next 5 to 10 minutes."[13]
"If you are vectored or directed by ATC out of the routine, MSA is a good guide to use to be at a safe and clear altitude from the terrain."
"MSAs should be included in approach and departure briefings with the intensions of how to use them."

I'm with you on MSA. What about terrain warning systems?
That's a good question since MSA factors into automated CFIT avoidance. For now, let's just talk about EGPWS, enhanced ground proximity warning systems. First, a word of caution:
"In today's high tech world it's really easy to become complacent."[14]
There's no question that pilots need to find ways to stay mentally active and involved in automated flight.

EGPWS. Terrain/obstacle warning. Briefly:
"With a Red 'pull up' warning, pull up immediately and climb to MSA (not when the warning stops)."
"Practice the pull up maneuver in the simulator; know the aircraft's capabilities and remember the feel of the aircraft."
"We practice CFIT scenarios in the simulator besides the classroom discussions."

Terrain warning systems have proven effective in reducing CFIT but, regarding the pull up maneuver:
"The evidence from incidents is that the aircraft is only maneuvered sufficient to stop the warning."
We discussed the need for personal standards earlier—the commitment to performing escape maneuvers is the same as that for go-arounds.

Situation awareness seems to have many facets.
Yes, there's much to appreciate. Now, before leaving the avionics topic, let me mention two other equipment applications. First, as you're aware, the non-precision approach is the ultimate challenge in a hazard-filled environment. One pilot recommends use of the radio altimeter (if one is installed) during non-precision approaches, and has established specific operating and readout guidelines. Although the box is far less sophisticated, the same pull up and escape commitment applies. Second, a helicopter pilot uses radar at low level "as a means of establishing a clear path ahead" and, in addition, "demands precision in maintaining radar altimeter heights." Just a reminder of an earlier point—know your systems and how to operate them effectively.

I can see that technology is a complex topic that requires more intensive discussion. Have we covered the CFIT spectrum?
Pretty much. But there is one other subject I'd like to close with that's not so procedurally oriented. After you've flow for a while you get a real feel for the airplane. It's as though you and your airplane have come together as a single unit functioning effortlessly. I'm sure you know the feeling, and there's a confident awareness associated with it. Many pilots also develop a feel for the flight ahead as they gather information and brief for it. They actually form mental pictures of what lies ahead—they visualize situations, places, events—and, in flight, use all their senses to "see" what's happening.

"Vision." Using all your senses. Pilots develop a heightened awareness by experiencing their environment, and express it this way:
"Situational awareness means referring to your surroundings to get the big picture."[15]

"For me, SA is knowing what your mission is and staying focused on it."[16]
"We need to use all of our senses to evaluate what's happening in the environment around us."[17]
"Situational awareness mandates you use your total senses to monitor flight parameters at all times."[18]
"I use all my senses to keep aware … you'll find your backside works great for identifying unusual vibrations or aircraft movements."[19]
"Situational awareness means being aware of your surroundings. [Clues] can be anything from avionics inputs to engine sounds, air noise, vibrations or lack thereof, and even aircraft smells"[20]
"An important part of SA is maintaining a good listening watch on the radio, especially in the terminal areas."[21]

And, as for taking action, trust your gut instincts. "The nagging feeling that something is not quite right is often unfailing in its precision. … If you get a gut feeling, respond to it, don't ignore it."[22] Good advice that wraps it up.

Thank you. You've helped me to gain a deeper understanding of—perhaps even a feel for—situation awareness and avoiding CFIT.

Postscript

The primary goal of this paper is to stress the value of pilot-to-pilot interaction and communication, particularly on the issue of near-ground operations.

Knowing that individual pilots have learned much from their training and experiences and developed personal rules of conduct that serve them, I conducted an informal CFIT avoidance survey with subscribers to Aviation.Org. Respondents are from many countries (including Turkey, Malaysia and The Netherlands) and have different experience levels and backgrounds. The results from a Professional Pilot survey added still another dimension.

Regulations, company policies, standardized procedures, etc. are necessary and desirable but, if we listen, those in the operating arena can fill in the gaps and provide added insight, not only to other pilots but to all in aviation.

Footnotes

[1]Pilot viewpoints on situation awareness are published in the Jan. and Feb. 2005 issues of Professional Pilot (abbr. Pro Pilot), 30 S. Quaker Lane, Suite 300, Alexandria VA 22314, in its letters section, Squawk Ident. Quotes from these letters receive attribution since they have appeared in print. Unless otherwise noted, other direct pilot quotes are the result of an informal survey of subscribers to the website aviation.org. Subscribers were promised anonymity and, therefore, these quotes are not attributed.
[2]John Taylor, Jr., NV. Pro Pilot, Jan. 2005, p14.
[3]Dan Gurney, UK. Aviation consultant. Private correspondence, Jan. 2005. Mr. Gurney made other important contributions to this paper.
[4]Flight Safety Foundation. CFIT Awareness and Prevention, 1995
[5]Jimmy Farris, MI. Pro Pilot, Feb. 2005, p22.
[6]Bill Howley, NJ. Professional Pilot, Jan. 2005, p16.
[7]Duane Tedesco, CT. Pro Pilot, Jan. 2005, p15.
[8]Robert Read, TX. Pro Pilot, Feb. 2005, p14.
[9]Christofer Dawson, NC. Pro Pilot, Feb. 2005, p18.
[10]CFR 14 Part 121.542.
[11]Bob and Skip Mudge, Stow, MA. Quantum-Pro cockpit management system.
[12]Steve Powell, TX. Pro Pilot, Feb. 2005, p18.
[13]Jeff Milam, AL. Pro Pilot, Feb. 2005, p18.
[14]Mike Miller, CA. Pro Pilot, Jan. 2005, p16.
[15]Jeffrey Miller, TX. Prof Pilot, Jan. 2005, p16.
[16]Greg van Liew, GA. Pro Pilot, Feb. 2005, p18.
[17]Laurence Wightman, TX. Pro Pilot, Feb. 2005, p18.
[18]Alan Rybarchyk, AZ. Pro Pilot, Jan. 2005, p14.
[19]Phillip Chastain, MO. Pro Pilot, Feb. 2005, p20.
[20]Bruce Haugsdal, MN. Pro Pilot, Feb. 2005, p22.
[21]Don Watkins, NM. Pro Pilot, Feb. 2005, p14.
[22]Eric Wickfield, MA. Situation Awareness | Losing the "Big Picture. Aviation Org website, March 1998.

THE ROLE OF WORKING MEMORY IN MAINTAINING SITUATION AWARENESS

Dr. Kamilla Run Johannsdottir
Department of Law and Social Sciences
University of Akureyri, Iceland

Dr. Chris M. Herdman
Centre for Applied Cognitive Research
Carleton University, Ottawa, Canada

Dr. Jo-Anne LeFevre
Centre for Applied Cognitive Research
Carleton University, Ottawa, Canada

Basic and applied research suggests that working memory (WM) supports situation awareness (SA) in dynamic environments. However, the relationship between WM and SA has not been well articulated. The present paper explores the potential role of WM in SA-based tasks by a) using a well-established WM model to conceptually link the two concepts and b) empirically testing this link. A dual-task paradigm was used where participants tracked an object against a moving background. Periodically, participants were required to either predict where the tracked object would be or to search for it. In addition to the tracking task participants concurrently performed one of four load tasks that separately taxed each of the four WM components (i.e. verbal, visual, spatial and central executive control). As predicted by the multi-component WM model (Baddeley, 1986; Logie, 1995) performing the SA tasks (prediction and search) relied on different WM subsystems. It is concluded that prediction involves the verbal subsystem whereas target search involves the spatial subsystem. The results support the role of WM in maintaining SA in a dynamic environment.

SA and WM

WM is the cognitive mechanism where information is integrated, manipulated and possibly recorded. Researchers have shown that the ability to activate and maintain sub-goals and intermediate solutions in WM is the key to success in many cognitive tasks. On this view, WM is believed to support the generation and maintenance of representations of complex task-environments. WM is what allows chess players to store sub–goals and to plan and anticipate future moves (Robbins et al., 1996). WM has also been shown to underlie the ability to solve problems (Carpenter, Just & Shell, 1990), and to engage in spatial reasoning (Shah & Miyake, 1996)

Given the role of WM in complex cognitive tasks, it seems likely that WM would be linked to SA. In accord with this view, based on fMRI measures, Perez et al. (2000) concluded that pilots' performance in a flight path maintenance task primarily involves cortical regions associated with WM. These cortical regions were more strongly differentiated for expert pilots compared to novice pilots. Perez et al. concluded that (a) WM supports pilots' comprehension of flight path information, including the anticipation of future actions and (b) the representation of information in WM becomes better defined with experience. Caretta, Perry, and Ree (1996) found that the ability to form and follow tactical plans and to communicate and interpret tactical information (e.g. threat prioritization) is related to spatial and verbal WM subsystems. Gugerty and Tirre (1995) found that maintaining awareness of location and avoiding hazards was highly correlated with WM measures.

In applied research, WM is frequently referred to as an important mechanism supporting SA (see Durso & Gronlund, 1999). However, the link between WM and SA has not been well articulated.

The goal of the present paper was to examine and strengthen the link between WM and SA. A multi-component model of WM (Baddley, 1986; Logie, 1995) was used as a theoretical framework.

The Multi-component Model of WM

The multi-component model of WM includes three subsystems for the maintenance of information, an episodic buffer for interaction with long-term memory (LTM) and an executive control system. Each subsystem uses different representational codes; visual, spatial, verbal. These subsystems are responsible for maintaining task-relevant information such as intermediate solutions and subgoals in order to carry out cognitive tasks. There may be other storage systems within WM for representing information in other modalities but the visual, spatial and the verbal systems are well established through research. Neuroimaging research, for example, has found support for distinct spatial and verbal WM systems (Smith, Jonides, & Koeppe, 1996).

The WM subsystems might play a critical role in maintaining SA in a dynamic task-environment. Research suggests that the verbal subsystem maintains cues (possibly linked to larger action plans in LTM) used to monitor and control action (Baddeley, Chincotta, & Adlam, 2001). The verbal subsystem might therefore be important for keeping track of and switching between multiple tasks. The verbal system has also been linked to the ability to make complex causal inferences during text comprehension (Shah & Miyake, 1996). The spatial subsystem is a movement-based system that involves planning and executing physical movements as well as representing the path between objects or target sequences (Quinn, 1991; Salway & Logie, 1995). The spatial subsystem is known to play an important role in tasks such as spatial reasoning (Shah & Miyake, 1996), chess playing (Robbins et al., 1996) and navigating (Garden, Cornoldi, & Logie, 2002). The visual subsystem has been less researched compared to the verbal and the spatial system. It refers to a temporary visual store for information such as shapes and colours.

The episodic buffer is a recent addition to the multi-component model and refers to a limited capacity buffer that represents coordinated information from the subsystems and from stored knowledge in LTM (Baddeley, 2000). The episodic buffer therefore is the connection between processing and representation in WM, and stored knowledge in LTM.

Finally, the central executive is a dedicated, possibly multi-dimensional, control system that is responsible for coordinating information from the various WM systems. The central executive also handles attention switching and controls both encoding and retrieval of information (Baddeley & Logie, 1999).

Overview of the Experiment

Based on the multiple-component model of WM, it was hypothesized that maintaining SA will differentially involve the WM subsystems depending on the specific SA demands. In particular, the verbal subsystem was assumed to support task switching and prediction based on prior knowledge about the task environment. The spatial subsystem was assumed to be important for representing spatial layout and executing movement. The central executive would be associated with coordinating tasks and selectively controlling attention.

A dual task paradigm was used in which a tracking task was combined with different WM load tasks. In the tracking task, participants tracked a target rectangle on a display by controlling a second rectangle with a mouse. Their task was to keep the controlled rectangle on top of the target rectangle. In addition to tracking, participants were also required to a) predict the future location of the tracked target and b) search for the tracked target. Periodically throughout the tracking task the tracked rectangle (pink) changed colour. A change to blue meant that the rectangle would disappear and then reappear in the lower right corner of the display. This is referred to as the prediction condition because the change from pink to blue was a consistent cue that would allow the participant to predict the future location of the tracked target. A change from pink to yellow meant that the tracked object would disappear and then reappear in one of the four corners of the display (randomly determined). This is referred to as the search condition because the change to yellow indicated that the target would appear in a corner: the participant was required to search the corners of the display to find the target.

The experiment therefore represented a task environment where fundamental aspects of SA were important. Specifically, participants were required to maintain a representation of task-relevant information in order to quickly activate predict or search activities. In addition, participants were required to use knowledge to predict the future location of the target and to coordinate the tracking task. The use of specific WM subsystems during tracking was assessed by introducing different WM load tasks that individually tapped into different subsystems of WM; verbal, visual, spatial and central executive.

Tracking was expected to depend primarily on the spatial WM subsystem. However, predicting and searching for the target object was assumed to rely on the verbal and the spatial system respectively. In order to predict or search for the tracked object, participants were required to maintain task-relevant information in WM (i.e., information regarding the colours and their meaning) that enables them to engage either a prediction or a search strategy. The spatial subsystem should play a strong role in the search condition because effective search requires a spatial representation of the display layout and an understanding of the relevant distances between the controlled rectangle and each corner of the display. This spatial representation would presumably enable the participant to quickly spot the target when it reappeared and importantly, to elicit the correct control input to quickly and accurately get to the display corner where the target appeared. Spatial layout and representation of movement have both been associated with the spatial subsystem (Quinn, 1991; Salway & Logie, 1995). It is expected, however, that predicting utilizes the verbal system. It is plausible to assume that the verbal system might be used for maintaining cues for activating relevant knowledge in LTM ("if blue go to the lower right corner").

In general, it was expected that responses to shifts in target location would be faster in the prediction than in the search condition. Undifferentiated response times between the prediction and search conditions would indicate that the participants were unable to use the predictive information and stored action plans. It was also expected that the introduction of the prediction and search conditions would involve the central executive system in addition to the subsystems. In order to coordinate task-related knowledge and switch from tracking to predicting or searching for the tracked object, the central executive must be involved.

Method

Participants

A total of 17 undergraduate students (10 females, 7 males) from Carleton University volunteered for this experiment. Participants received course credit for their participation.

Apparatus, Stimuli, Design, and Procedure

The experiment was controlled using E-prime software. The visual display was presented on a 17-inch SVGA colour monitor. To provide enhanced realism, a map of the greater-Ottawa area was presented as background on the display. This map was shown in grey colour/tones and moved vertically up the screen at a steady rate.

The tracking task included two stimuli: a pink target rectangle and a red controlled rectangle, 1cm x 2cm in size. These stimuli were superimposed over the moving map: the red and pink colours were easily distinguished from the background and from each other. The pink rectangle moved along both x-and y-axis according to a pre-defined loop which was independent of the background map movement. The red rectangle was controlled by the participant using a mouse: participants were instructed to use the mouse to keep the red rectangle on the top of the pink rectangle. Data for the X and Y position of the two rectangles was collected at 10 Hz.

Participants performed the primary tracking task combined with one of 4 load tasks: discrimination of shapes (visual), rhyming (verbal), tapping a defined pattern on a keypad (spatial) and tapping randomly on a keypad (central executive). The tracking and load tasks were performed alone (single-task) and together (dual-task conditions). Hence, there were nine conditions for the experiment. Each condition consisted of ten 30-second trials. The nine conditions were presented in a random order to each participant.

Approximately twice in each 30-second trial, the target rectangle changed colour. A change from pink to blue indicated that the rectangle would reappear in the lower right corner (prediction). A change from pink to yellow indicated that the rectangle would reappear in one of the four corners (search). Participants were instructed to move (as quickly as possible) the tracking (controlled) rectangle to the corner where the rectangle reappeared. They then followed the (blue or yellow) rectangle as it joined the defined tracking loop at which time the rectangle became pink again and the normal tracking task continued for the remainder of the trial or until next colour change occurred. The time it took participants to move the controlled rectangle to the corner of the display where the tracked rectangle reappeared was measured.

Results

An alpha level of .05 was adopted throughout this research. For comparisons between individual conditions, 95% confidence intervals were used (Loftus & Masson, 1994). Rather than recalculating the difference between conditions based on planned comparison, confidence intervals allow for a simple (visual) heuristics where a difference between two conditions is judged to be significant when the confidence intervals of two conditions overlap by ¼ or less of the total interval. A significant difference between two conditions is also referred to as a critical difference and can be calculated by multiplying the confidence interval by the square root of 2.

RMSE Tracking

RMSE tracking was analyzed in a one-way repeated-measures ANOVA of load task (baseline, verbal, visual, spatial and central executive). A significant effect of load task on tracking, $F(4, 64) = 14.917$, $MSE = 5.051$, indicated that tracking was generally worse in the dual-task conditions than in the single-task tracking condition. More importantly, the spatial task resulted in significantly larger tracking RMSE as compared to the single-task condition as well as the verbal and the visual conditions. This provides support for the notion that tracking involves the spatial, movement-based subsystem of WM. The central executive task (random tapping) resulted in similar decrements in tracking performance to those in the spatial task condition, suggesting that tracking did not involve additional central executive resources.

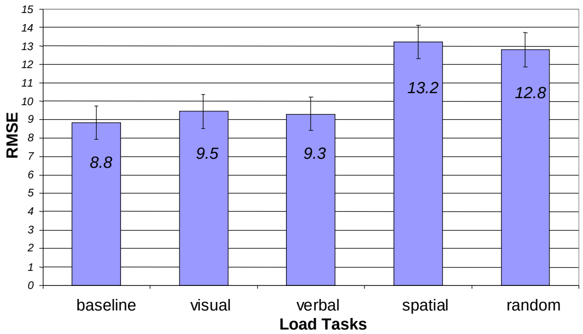


Figure 1. *RMSE for baseline (tracking only) and tracking with the different load tasks.*

Prediction and Search

The time it took participants to move the controlled rectangle to the correct corner where the tracked rectangle reappeared was analyzed in 2 (task: prediction vs. search) by 5 (condition: single-task, dual-task visual, verbal, spatial and random) repeated-measures ANOVA. As expected, there was a significant main effect of task, $F(1, 15) = 58.44$, $MSE = 39805.53$, and of condition, $F(4, 60) = 2.72$, $MSE = 10869.96$. The analysis also revealed a significant task by condition interaction $F(4, 60) = 7.24$, $MSE = 8511.12$.

The 95% confidence intervals (see Figure 2) show that participants were quicker to respond when they could predict the location of target rectangle as compared to when they had to search for the target. This shows that participants were able to use knowledge to predict target location.

Of primary interest was that the verbal task caused significantly more impairment in predicting the target location than the spatial task. This supports the hypothesis that predicting involves the verbal subsystem in WM.

For search, only the spatial task resulted in significant impairments relative to the single-task condition. This supports the hypothesis that searching for a target location depends on the spatial subsystem in WM.

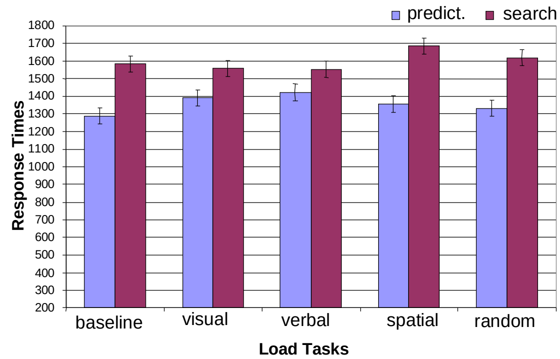


Figure 2. *RTs for moving the controlled rectangle to the corner where the tracked rectangle reappeared.*

In sum, the results show that prediction of a target location involves mainly the verbal WM subsystem whereas search for a target location primarily involves the spatial subsystem. The central executive control system did not play a particular role in the SA based tasks tested here. Analysis of the load tasks revealed increased error in random tapping (central executive task) while participants were engaged in search or prediction. The increased error might indicate that participants shed the central executive task in order to engage in search or prediction suggesting a role of the central executive system in coordinating the tasks and selectively controlling attention.

Conclusion

The present experiment supports the notion that WM subsystems differentially support SA in complex task environments. Specifically, it was shown that the verbal WM subsystem supports SA related to the prediction of target location, whereas the spatial subsystem supports SA related to target search.

Previous research has suggested a role for the spatial subsystem in maintaining navigational awareness (Arez, 1991; Gugerty, 1997). The unique contribution of the present experiment is in showing that the verbal subsystem also supports specific aspect of SA. By maintaining active cues in the verbal subsystem, participants could quickly retrieve the task-relevant knowledge (e.g., if blue move to the lower right corner) to predict target location.

The results provide support to the literature which has suggested a complex role of WM in supporting SA. For example, Caretta et al. (1996) found correlations between SA and spatial and verbal WM tasks as well as between SA and spatial reasoning tasks. Gugerty and Tirre (1995) reported correlations between SA for surrounding traffic in a driving simulation and various WM measures. Similarly, Aretz (1991) found that when pilots had to control a simulated aircraft and perform a difficult navigation task they switched from using spatial WM to using verbal WM. The present study extends those findings by showing that SA is supported by maintaining and updating information in both the verbal and the spatial subsystems. Further studies are needed to better understand the link between SA and the various WM systems, in particular the central executive control system and the episodic buffer.

Few attempts have been made to connect SA to underlying cognitive mechanisms. One reason for this is that SA is commonly viewed as a process or representation that an operator can consciously introspect upon (Endsley, 1995). Accordingly, researchers have often used conscious reports as a primary measure of SA and the cognitive mechanisms that are fundamental to generating and maintaining SA have been of little interest. The present research, however, suggests that SA can be conceptualized in terms of specific WM subsystems.

References

Aretz, A. J. (1991). The design of electronic map display. *Human Factors*, *33*, 85-101.

Baddeley, A. D. (1986). *Working Memory*. Oxford: Clarendon Press.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Science*, *4*, 417-423.

Baddeley, A. D., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, *130*, 641-657.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), Models of Working Memory (pp. 28-61). New York: Cambridge University Press.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the raven progressive matrices test. *Psychological Review*, *97*, 404-431.

Carretta, T. R., Perry, D. C., & Ree, M. J. (1996). Prediction of situational awareness in F-15 pilots. *The International Journal of Aviation Psychology*, *6*, 21-41.

Durso, F. T. & Gronlund, S. D. (1999) Situation Awareness. In T. F. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay and M. T. H. Chi (Eds.). *Handbook of Applied Cognition*, (pp. 283-313).

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*, 65-84.

Garden, S., Cornoldi, C., & Logie, R. H. (2002). Visuo-spatial working memory in navigation. *Applied cognitive psychology*, *16*, 35-50.

Gugerty, L. C. (1997). Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory. *Journal of Experimental Psychology: Applied*, *3*, 1-26.

Gugerty, L. J. & Tirre, W. C. (1995). Comparing explicit and implicit measures of situation awareness. In D. J. Garland & M. R. Endsley, (Eds.), *Experimental Analysis and Measurements in Situation Awareness*. Daytona Beach: Embry-Riddle Aeronautical University Press.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476-490.

Logie, R. H. (1995). *Visuo-Spatial Working Memory*. Hove, UK: Erlbaum.

Peres, M., Van De Moortele, P. F., Pierard, C., Pharm, D., Lehericy, S., Satabin, P., Le Bihan, D., & Guezennec, C. (2000). Functional magnetic resonance imaging of mental strategy in a simulated aviation performance task. *Aviation, Space and Environmental Medicine*, *71*, 1218-1231.

Robbins, T. W., Anderson, E. J., Barker, D. R., Bradley, A. C., Fearnyhough, C., Henson, R., & Hudson, S. R. (1996). Working memory in chess. *Memory & Cognition*, *24*, 83-93.

Salway, A. F. S., & Logie, R. H. (1995). Visuospatial working memory, movement control and executive demands. *British Journal of Psychology*, *86*, 253-269.

Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cognition*, *65*, 167-196.

Acknowledgements

This research was supported by the National Sciences and Engineering Research Council of Canada (NSERC), the Centre for Research in Earth and Space Technology (CRESTech), and by DND Canada through grants to Dr. C. Herdman. We thank C. Michalyna for programming and technical assistance on this project.

A SIMULATION EVALUATION OF A HUMAN-CENTERED APPROACH TO FLIGHT DECK PROCEDURES AND AUTOMATION FOR EN ROUTE FREE MANEUVERING

Walter W. Johnson and Vernol Battiste
NASA Ames Research Center
Moffett Field, CA

Stacie Granada, **Nancy Johnson**, and **Arik Quang Dao**
San Jose State University
Moffett Field, CA

Dominic Wong and **Anthony Tang**
QSS Group Inc.
Moffett Field, CA

A joint simulation was conducted by NASA Ames and Langley Research Centers. This paper presents flight deck performance and subjective data collected at the NASA Ames Research Center. During the simulation of en route free maneuvering, the presence and mixture of managed and autonomous aircraft was manipulated, as was the number of autonomous aircraft.. These manipulations allowed for an examination of the viability of both conducting mixed AFR-IFR operations, and of substantially increasing en route traffic through insertion of AFR aircraft. The Ames airside performance and pilot comment data support the safety and feasibility of the concept, with double en route capacity appearing to be easily achievable. This work was supported by NASA's Advanced Air Transportation Technologies (AATT) project.

Introduction

In June 2004, research teams at the NASA Ames and Langley Research Centers conducted a joint human-in-the-loop simulation investigating the feasibility and operational benefits of a free flight concept under consideration by NASA's Distributed Air Ground Traffic Management (DAG-TM) effort. The goal of DAG-TM is the expansion of airspace capacity and, to this end, several concepts have been developed and evaluated as part of this effort. The concept evaluated in this simulation was called En Route Free Maneuvering. The premise was that greater efficiency and capacity could be gained through a redistribution of roles and responsibilities, and attendant decision-making, to the aircraft operators (e.g., flight crews) and the air traffic management system (e.g., air traffic controllers). The envisioned solution requires new human-centered operational paradigms enabled by advances in decision support tools: information sharing; communication, navigation, and surveillance; air traffic management technologies; and procedures supporting distributed separation responsibilities.

Simulation Overview

The simulation investigated the feasibility of an en route free maneuvering concept with respect to traffic scalability and airspace mixture. The free maneuvering concept assigned the en route autonomous aircraft two tasks. First, autonomous aircraft had the task of maintaining separation from other aircraft. Second, autonomous aircraft arrivals were responsible for arriving at a TRACON meter fix at a required time of arrival. Scalability refers to the ability to increase the density of en route aircraft by adding autonomous aircraft, while airspace mixture refers to the mixing of autonomous and managed aircraft in the same airspace. Autonomous aircraft are aircraft flying according to Autonomous Flight Rules (AFR) designed specifically for free maneuvering aircraft, while managed aircraft are aircraft flying according to standard Instrument Flight Rules (IFR) by which aircraft fly today. The simulation was manned by a team of certified professional controllers and commercial-rated pilots who operated within a simulated airspace similar to the northwest portion of the Fort Worth Air Route Traffic Control Center (ZFW). Traffic scenarios were created to simulate realistic traffic flow into and out of the Dallas-Ft.Worth (DFW) airport, and also overflights through the ZFW airspace. Varying traffic volume between scenarios was accomplished by altering only the number of overflights, and thus Scalability was manipulated only for en route (non-arrival) flight. The arrival problem, while demanding, remained relatively constant throughout all scenarios, but was impacted by the increased overflights passing near and around the arrival stream.

The DAG-TM simulation environment was distributed across two NASA facilities and several laboratories. At the Ames Research Center, the Airspace Operations Laboratory (AOL) provided aircraft target generation and ran the human-in-the-loop air traffic control (ATC) part of the simulation. The AOL utilized professional air traffic controllers and also provided "pseudopilots," who were given specialized tools that allowed them to simultaneously control multiple background IFR aircraft. The use of IFR pseudopilots was necessary because the volume of traffic was such that it was impossible to assign one person for every aircraft in the airspace.

AFR human-in-the-loop flight deck simulation and evaluations were conducted at the NASA Ames Flight Deck Display Research Laboratory and the NASA Air Traffic Operations Laboratory. Both centers were responsible for gathering data on professional pilots serving as participants. Also, as with the AOL, these laboratories provided pseudopilots to control background AFR traffic.

The Ames and Langley AFR flight decks and pseudopilot stations were different and flew in separate areas of the airspace. Only the Ames AFR flight decks are examined in this report.

The entire scope of the DAG-TM En Route Free Maneuvering Simulation is far too large to describe in the present conference report. Those interested in greater detail are referred to the Joint NASA Ames/Langley Joint Simulation Final Report (Raytheon ATMSDI Team, 2004). The present report focuses solely on selected aspects the NASA Ames Flight Deck simulation evaluations.

Roles and Responsibilities

The DAG En Route Free Maneuvering concept placed a large number of new responsibilities on the AFR pilots, responsibilities that have previously been reserved to air traffic control. First, and foremost, the AFR pilots were assigned the responsibility for maintaining separation from all IFR (ATC managed) aircraft. In addition, they were responsible for maintaining separation from other AFR aircraft according to a set of "rules-of-the-road" (c.f., Johnson, Canton, Battiste, & Johnson, 2005). When responsible, or "burdened," AFR pilots were required to resolve any impending loss-of-separation (LOS) prior to two minutes to the predicted time of the LOS. In addition, when attempting to resolve a conflict, AFR pilots were not allowed to create a predicted LOS that was less than four minutes in the future. Some AFR flights were solely overflights, but others flew arrivals. These AFR arrivals were given the responsibility for meeting a required-time-of-arrival (RTA) (+/- 15s) for the DFW Meter Fix BAMBE. The arrivals began the simulation during cruise ~200 NM from BAMBE, and were assigned their RTA when they were approximately 160 nm (~20 min) from BAMBE. In addition to the RTA restriction, AFR aircraft were required to cross BAMBE at 11000 ft (+/- 300 ft) and 250 kts (+/- 10 kts). If these crossing restrictions could not be met, then the TRACON controller could refuse the AFR aircraft permission to cross into the TRACON, or, at the controller's discretion, new crossing restrictions could be given. At BAMBE, the status of aircraft transitioned back to that of an IFR aircraft.

NASA Ames Flight Deck Implementation

The NASA Ames flight deck implementation for the DAG-TM En Route Free Maneuvering work is based on the Airborne Management of En route Separation Display (AMES Display). The AMES Display was developed between 1995 and 2000 to support initial examinations of Free Flight by NASA's Advanced Aeronautical Transportation Technologies Project (Johnson, Battiste, Delzell, Holland, Belcher, & Jordan, 1997; Johnson, Battiste, & Holland, 1999). The AMES Display was not only a display, but was an integrated approach to the display, manipulation, and management of flight path information required for free maneuvering in the en route environment. As such, a free maneuvering concept was actually embedded in its design, although the present DAG-TM work considerably surpasses this initial work in scope and requirements. Nevertheless, important design decisions made during the development of the AMES Display continued to inform the DAG-TM work. Perhaps the most important of these concerned the role of intent information.

Intent Information: Free maneuvering responsibilities could not be met using today's flight deck resources. Specifically, conflict detection and resolution (CD&R) capabilities are required. In order to predict the "when" and "where" of conflicts, it is necessary to predict the flight paths of the surrounding aircraft, and this, in turn, requires information about the future flight path. There are two general approaches to this problem. One approach bases conflict predictions on the broadcast of aircraft state information (current position and 3-D velocity), or state information plus Mode Control Panel (MCP) commanded intent (in this case, the location at which an aircraft automatically levels off during a climb or descent also is broadcast). In other words, this first approach bases predictions solely on current-state variables. However, in order to insure detection of conflicts prior to a minimum time to LOS, this approach also requires restricting an aircraft from changing its flight path (i.e., state) until the pilot assures that it will not create a near-term conflict with other aircraft. If a pilot was trying to follow a flight plan using this approach, it would require the flight crew to monitor and ensure the safety of maneuvers at every trajectory change point. This could substantially increase the mental workload of the flight crew. The second general approach is to share intent information based on flight plans (e.g., broadcasting state plus future trajectory change points in the flight plan). Recently, the broadcast of partial flight plan information, in the form of the next four trajectory change points, has been embraced by the RTCA (2002). However, if flight intent information is represented by a finite number of trajectory change

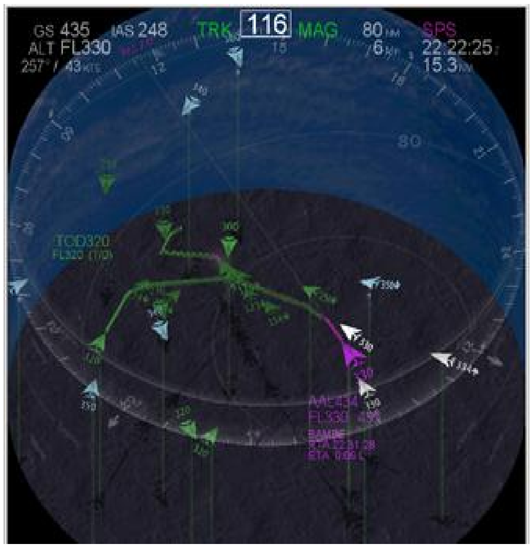


Figure 1. *3D CSD. To view pictures in color or to download a demo (http://human-factors.arc. nasa. gov/ihh/cdti/ download.html).*

points (e.g., as in the four points suggested by the RTCA), the time and area represented by the information could vary greatly (from representing a great amount of time/airspace to representing very little amount of time/airspace). Thus new conflicts could appear with little warning. Therefore, a different approach was utilized in the development of the AMES Display. Specifically, the AMES Display uses information regarding the *entire* flight plan. By incorporating the entire flight plan, the information presented on the AMES Display ensures (at the design level) that the flight crew has an adequate look-ahead time and eliminates abbreviated flight plans as the cause of alerts that "pop-up" with insufficient time to resolve.

Flight Deck Tools: The tools that supported CD&R, along with pilot situation awareness, were all integrated into a PC-based Cockpit Situation Display (CSD) shown in Figure 1. This display was built on the AMES Display and should be considered an extension of that display. Many of these tools were also integrated into an aircraft simulator based on the Multi-Aircraft Simulator (MACS) developed by the NASA Ames Airspace Operations Laboratory (Raytheon ATMSDI Team, 2004).

The CSD was a 3D perspective display that allowed the pilot to display the 4D flight plans of traffic, see traffic conflicts, and manipulate the viewing angle. Graphical path replanning capabilities of the 3D CSD were integrated with the Flight Management System (FMS), such that a pilot could graphically design a conflict-free 3D route, then load and execute it from the CSD interface. Thus all conflict resolutions and replanned flight paths depended on these human-centered tools (see Canton, Refai, Johnson, & Battiste, 2005, for a fuller discussion of the CD&R implementation). In addition, the CSD also incorporated a tool for the management of RTAs that allowed the pilot to enter, execute, and monitor RTA conformance, all from the CSD. This display is described more fully by Granada, Dao, Wong, Johnson, and Battiste (2005), and in a users manual that can be found at: http://human-factors.arc.nasa.gov/ihh/cdti/download.html.

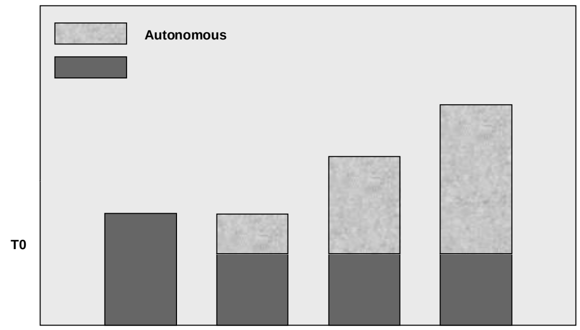
The MACS simulator emulated a Boeing 757, and included an FMS emulation. The pilot was able to use the FMS in the standard fashion to design new routes, but these would show up on the CSD and be tested for conflicts. If the CSD showed them to be conflict free, the pilot could then execute these routes from the FMS. In addition to the FMS capabilities, the pilots could take their aircraft off of the flight plan and onto a 3D vector using the MCP. Again, the vector would be probed for conflicts, with the pilot executing the vector only if there was at least four minutes of clear path ahead of the aircraft. Finally, the system allowed the pilots to use the FMS to recapture their flight plan by using the FMS to take them directly to any subsequent waypoint in their flight plan.

Method

Experimental Design

The experimental design was a comparison of four conditions (Figure 2). These conditions compared four levels of AFR/IFR Scalability/Mixture in the en route ZFW Center airspace. "T0" represents a traffic threshold that approximates the current-day monitor alert parameter for the simulated airspace. "T1" is a projected threshold above which managed-only operations cannot be achieved. T0 and T1 levels were determined in a prior controller-in-the-loop study at NASA Ames. "C1" represents a high IFR traffic volume that can be handled using normal ATC operations. "C2" is the same traffic volume but with 25% of the aircraft AFR, and 75% IFR. "C3" maintained the same number of IFR aircraft as in "C2", but added AFR aircraft to increase the total traffic by an average of 60%, while in "C4" this increase was approximately 100%. While these factors changed the density and mix of traffic in the en route airspace, the rate/mix of arrivals into DFW remained constant, with 20%-25% of them being AFR in conditions C2-C4.

This design allows two critical comparisons. First, by comparing C1 and C2 we can determine if introducing mixed operations alone enhances or degrades performance and workload/acceptability. Second, comparisons of C2-C4 test scalability. That is, can AFR aircraft be used to increase traffic load by approximately 50% and 100% without significantly degrading overall operations?



Participants

Ten current or retired airline pilots flew nine Ames experimental aircraft in the simulations. Eight flew desktop single-aircraft MACS flight deck stations, and two served as captain and first officer in the Advanced Concepts Flight Simulator (ACFS), a high fidelity reconfigurable simulator with full window visuals. The total number of flight hours for these pilots ranged from 4,500 to 22,000, with a mean of approximately 11,000 hours. All the Ames pilots had glass cockpit experience ranging from 85 to 8,000 hours, with a group mean of approximately 4,000 total hours. All participants had previous experience with the DAG-TM project, having participated in several previous studies.

Procedure

The simulation was conducted over eight days. The first two days were used for orientation and final training. Initial pilot training had taken place during two previous dry run and dress rehearsal simulations. Four simulation scenarios per day were conducted on days 3-7, with day seven reserved to replace scenarios if there were unanticipated problems (there were none). The eighth day was used for extensive debriefing. On each of the days 3-6, all of the four conditions (C1-C4) were presented. Across these days, all eight single station pilots flew two arrivals and two overflights for each of the four conditions. The two pilots in the ACFS Simulator always flew arrivals. For half of the scenarios, the Ames aircraft began in the ZFW Ardmore airspace, while on the other half they began in the ZFW Amarillo High/Wichita Falls High airspace.

The overflights had no special requirements other than to maintain separation from other aircraft. All of the arrivals began between 160 NM and 200 NM from the BAMBE arrival meter fix. As the arrivals approached 160NM from BAMBE, they were given RTAs by ATC. These RTAs were designed to require a substantial delay relative to the aircraft's present estimated time of arrival (ETA) at BAMBE. This delay was large enough that the pilots could not accommodate it through a speed change alone, but required the pilots to modify their flight plans by "stretching" them using their path replanning tools. The arrival aircraft were at various en route altitudes above FL300, and reached their top of descent approximately 80 NM from BAMBE. Data collection ceased for an aircraft when the aircraft reached the BAMBE meter fix (at 11000 ft).

Results and Discussion

Single Pilot AFR Meter Fix Conformance: There was no significant difference in objectively measured performance as a function of any of the conditions. Therefore, from the Ames flight deck perspective, neither mixed operations, nor increased en route traffic load had any effect on the ability of the arriving flights to meet the meter fix constraints. Among the AFR flights, one failed to meet the speed constraint, and one failed to meet the altitude constraint, but all met the RTA constraint.

Single Pilot AFR Self-Separation Performance: There were no LOS incidents for the Ames AFR aircraft. There were 139 conflicts resolved. Of these, 122 were detected prior to four minutes to conflict. The concept requires that AFR aircraft detect conflicts at least four minutes prior to the expected LOS, that no aircraft maneuver such that they create conflicts with expected losses of separation under four minutes, and that the AFR aircraft resolve conflicts prior to two minutes to LOS. Figure 3 shows 17 late alerts (less than four minutes to LOS). Consistent with the increased traffic load, the majority (11 late conflicts) occurred in the C4 condition. All 17 were due to an aircraft executing a maneuver which brought about the late alert. However, a maneuver by an Ames single piloted AFR aircraft was responsible for these late conflicts (concept violations) in only four instances. Furthermore, in three of the four cases, this was due to a flaw in the conflict resolution software (indicating to the pilot that the maneuver was conflict free, when it was not). The reason for the remaining apparent procedural error remains to be determined.

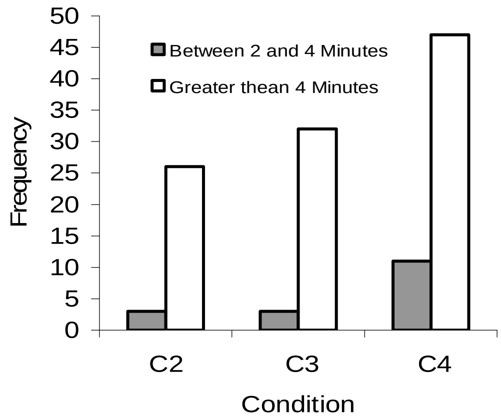


Figure 3. *Number of conflicts detected 2-4 minutes to LOS, and beyond 4 minutes to LOS*

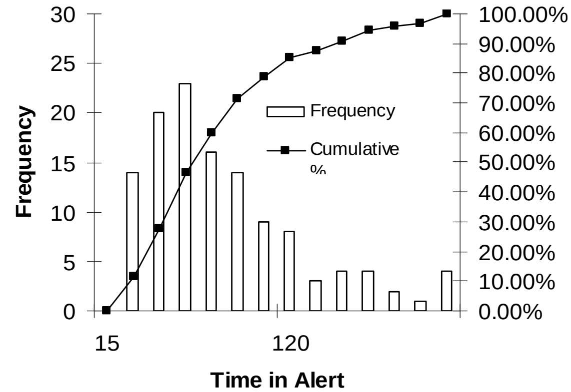


Figure 5. *Distribution of times required to resolve conflict alerts*

Figure 4 shows a total of four conflicts with late resolutions (under two minutes to LOS), with three of them due to a software flaw in which the conflict detection did not function properly. The fourth instance was associated with the aforementioned incident where the pilot maneuvered into a conflict at some point between two and four minutes to LOS, with the result being that the pilot was not able to resolve the conflict until there was less than two minutes until the projected LOS.

Figure 5 shows the amount of time needed to resolve conflicts. While not an integral part of the concept, the Ames AFR pilots were asked to resolve all conflicts within two minutes (120 sec) of receiving a conflict alert. Figure 5 shows that this criterion was met approximately 80% of the time.

Finally, Figure 6 shows the number of times Ownship and Intruder resolved the conflict as a function of burdening. While the IFR aircraft almost never resolved a conflict (it was never burdened), it is noteworthy that almost 1/3 of the non-burdened aircraft resolved the conflicts in AFR-AFR conflicts.

Subjective Assessments: The subjective assessment of pilot workload was measured following each run using the Modified Cooper Harper (MCH) workload scale. The MCH allows for ratings between 1 (Very easy/workload insignificant) and 10 (Impossible/task abandoned, unable to apply sufficient effort). Pilot responses across all simulation trials ranged from 1 to 6. However, approximately 98 percent of responses ranged from 1 to 3. In order to receive a rating between 1 and 3, it must be possible to complete the task, and workload must be perceived as tolerable and satisfactory. Ratings from 4 to 6 suggest that task workload is high but not high enough to impact performance on the primary task.

Figure 7 shows the average workload ratings of arrivals and overflights in each condition (±1 SD). Not surprisingly, workload ratings were higher for the arrival flights than for overflights in all conditions. Pilots also responded unanimously with the lowest

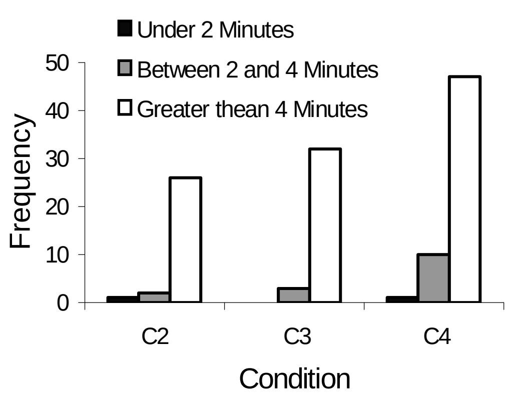


Figure 4. *Number of conflicts resolved prior to 4 minutes to LOS, 2 minutes to LOS, & under 2 minutes to LOS, as a function of condition*

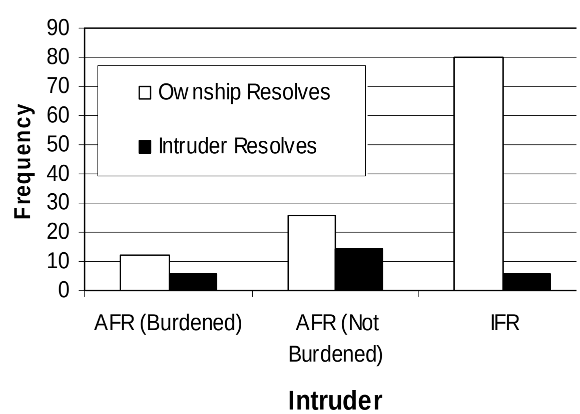


Figure 6. *Number of conflict resolutions for Ownship and Intruder as a function of burdening*

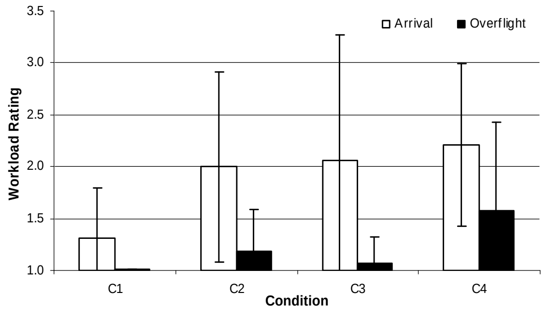


Figure 7. *MCH workload ratings as a function of flight phase and condition*

possible workload rating (1) for the managed overflight runs (C1). In addition, Figure 7 shows an increase in perceived workload from all managed (C1) to mixed (C2-C4) operations for both flight types, but workload remained acceptably low.

Nine of the ten pilots completed post-simulation questionnaires that elicited their opinion of the viability of the DAG-TM Free Flight concept and of the implementation. Figure 8 shows that the majority preferred AFR (Free Flight) for Safety, Workload, Ease of Meeting RTA, and Situation Awareness.

Conclusions

The results support the viability of the En Route Free Maneuvering concept. Not only were the concept requirements met, but the subjective data show that the implementation resulted in low workload, and a high degree of pilot acceptance. While these results must be considered as very preliminary, especially given the near perfect nature of the information exchanged between aircraft. Also, the lack of disruptive events (e.g., weather and emergencies) also limits the generality of the study.

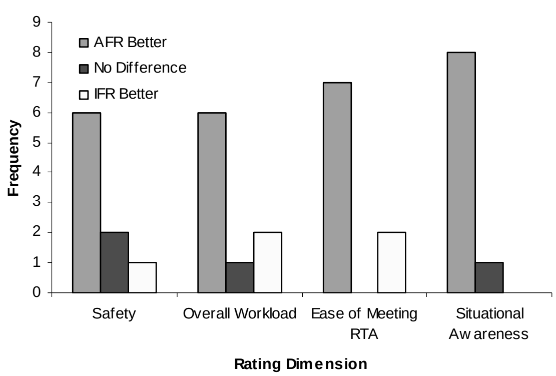


Figure 8. *Number of pilots prefering AFR or IFR flight rules as a function of safety, workload, RTA, and situational awareness*

That said, the excellent performance of the pilots, along with the low workload, suggests that the design of the tools and procedures was very successful, both in terms of software design and human factors design. The high ratings given to situation awareness and ease in meeting RTA also bear this out. That this was achieved while intimately involving the pilot in the resolutions of conflicts, and in other critical decision-making, shows the viability of a human-centered approach to free flight.

References

Canton, R., Refai, M., Johnson, W. W., Battiste, V. (2005). Development and integration of human-centered conflict display and resolution tools for airborne autonomous operations. *Submission to the 2005 International Symposium on Aviation Psychology*, Oklahoma City, OK

Granada, S., Dao, A., Wong, D., Johnson, W.W., Battiste, V. (2005). Development and integration of a human-centered volumetric cockpit situational display for distributed air-ground operations. *Submission to the 2005 International Symposium on Aviation Psychology*, Oklahoma City, OK

Johnson, N. H., Canton, R., Battiste, V., and Johnson, W. W. (2005). Rules of the Road: Requirements and Implementation for a Simulation of En-route Self-Separation. *Submission to the 2005 International Symposium on Aviation Psychology*, Oklahoma City, OK

Johnson, W. W., Battiste, V., and Holland, S. (1999). A cockpit display designed to enable limited flight deck separation responsibility. *Proceedings of the 1999 World Aviation Conference*, Anaheim, CA.

Johnson, W., Battiste, V., Delzell, S., Holland, S., Belcher, S., and Jordan, K. (1997). Development and demonstration of a prototype free flight CDTI. *Proceedings of the SAE/AIAA World Aviation Congress*, Anaheim, CA.

RTCA. (1995). *Free flight implementation* (RTCA Task-Force 3 Report). Washington, D.C.: RTCA Inc.

RTCA. (2002). *Minimum Aviation System Performance Standards for Automatic Dependent Surveillance Broadcast*. RTCA/DO-242A. RTCA, Inc. Washington, DC.

Raytheon ATMSDI Team. (2004). *Joint NASA Ames/Langley DAG-TM Simulation Final Report* NASA Ames Research Center. Moffett Field, CA.

DISTRIBUTED AIR / GROUND TRAFFIC MANAGEMENT EN ROUTE FREE MANEUVERING RULES OF THE ROAD: REQUIREMENTS AND IMPLEMENTATION FOR A SIMULATION OF EN-ROUTE SELF-SEPARATION

Nancy H. Johnson
San Jose State University
NASA Ames Research Center

Riva Canton
Raytheon-QSS Corporation
NASA Ames Research Center

Vernol Battiste
Walter Johnson
NASA Ames Research Center

National Aeronautics and Space Administration's (NASA) Distributed Air Ground Traffic Management (DAG-TM) program has recently investigated a concept called "En-route Free Maneuvering" as a proposed solution for expanding airspace capacity limits. A critical element for this concept is conflict detection and resolution (CD&R) using the 3D cockpit situation display (CSD). The only fielded system performing some of this function is the Traffic Alert & Collision Avoidance System (TCAS), a radar-based alerting system used by most commercial aircraft for collision detection and avoidance. TCAS is inappropriate for an en-route self-separation application due to its reactive nature, and inherent lack of flexibility. Therefore, a new system was designed with improved intent information in the form of 4D flight plans, broadcast and shared amongst en-route aircraft, which in turn allowed these aircraft to detect and resolve conflicts well in advance of a projected conflict. A key element in this approach is ensuring that *burdening*, the assignment of final responsibility for conflict resolution is clearly assigned to the aircraft not in right-of-way. The basis for this burdening is called the *rules-of-the-road* (ROR), a term taken from the rules designed for guiding collision avoidance in VFR (visual flight rules). Given the potential complexity of determining burdening assignment, the automation described herein computes assignment using these rules, and then notifies the crew if it has the right-of-way or is burdened to resolve the conflict.

Introduction

DAG-TM is a proposed solution for expanding airspace capacity limits. It alters the roles and responsibilities of the stakeholders – airlines and air-traffic control - to permit more user-preferred routing, increased flexibility, increased system capacity, and improved operational efficiency. It is based on the fundamental premise that the National Airspace System (NAS) participants can be information suppliers and users, thereby enabling collaboration at all levels of traffic management decision-making. The success of this proposed, future environment may depend greatly on new human-centered operational paradigms enabled by technological and procedural innovations (Raytheon ATMSDI Team. 2004).

Air travel has advanced from an uncontrolled "see and avoid" environment, to vastly increased numbers of aircraft tightly controlled by ground facilities. As it moves on into the next generation, one that supports an increase in capacity and free flight, distributed control is suggested as a viable air traffic management (ATM) model. Distributed control refers to the delegation of responsibility between the air traffic service provider (ATSP) and the flight crew, defining whom maintains separation assurance. In this environment, the old rules-of-the-road that supported aircraft in a visual-only airspace will no longer work effectively to ensure safety of flight. A new set of rules for 'autonomous aircraft' that share separation responsibilities in a free flight environment are required.

Current Operations

In current operations, often-used flight rules are virtually second nature to the pilot who is being managed by air traffic control. Pilots flying commercial class aircraft are less likely to need to refer to the visual flight rules although there are rules that become ingrained much in the same manner as a driver interacts with traffic laws while driving a car. Specifically, visual flight rules are a set of regulations that a pilot may operate under when weather conditions meet certain minimum requirements. Under VFR, the pilot controls the attitude of the aircraft by relying on what can be seen out the window, although this may be supplemented by referring to the instrument panel. A pilot flying under VFR is usually required to stay a specified distance away from cloud formations and remain in areas where the visibility meets minimum requirements. In VFR, the pilot is responsible for seeing and avoiding other aircraft, terrain, and obstructions such as buildings and towers. Being in contact with air traffic control is optional in most airspace, and the pilot is usually allowed to select the course and altitude to be flown based on VFR direction of flight and altitude rules (http://www.fact-index.com). The pilot may also choose to navigate by reference to visual landmarks and/or utilize electronic navigation aids. In a distributed control environment, the pilot would also be responsible for maintaining separation from other aircraft.

Current commercial transport operations utilize a collision avoidance system called Traffic Alert & Collision Avoidance System (TCAS). TCAS scans radar information of proximal traffic to determine distance and closure rate. If TCAS detects that an aircraft's distance and closure rate are potentially threatening to Ownship, it will generate a traffic advisory (TA) or a resolution advisory (RA) to the crew. Both advisories are displayed on the TCAS display screen and are accompanied by an auditory alert. If necessary, TCAS will compute a pitch command to avoid collision. At this time, TCAS is limited to vertical guidance and cannot coordinate aircraft performance standards into the resolution advisories. With TCAS II, the pitch commands are coordinated with the other conflicting aircraft – up to three **-** to avoid escape maneuvers in the same direction.

An inherent problem with TCAS is that it is reactive and involves little planning on the part of the pilot. A resolution advisory provides the pilot with a 25-second response time before loss of separation (LOS). TCAS logic does not incorporate flight path intent and as a result, when crews respond to a TCAS RA, they are instructed to perform either a vertical maneuver or to remain at current altitude. A suggested vertical maneuver may send the conflicting aircraft off their intended flight path and unexpectedly into another ATC sector. And, because TCAS does not have flight path information, false alerts are frequent in busy, high workload terminal areas.

The TCAS system has been a tremendous asset, however, the free flight environment may be one in which the most effective and efficient resolutions are based on *planned* maneuvers, which require information about aircraft intent. Therefore, in the least, TCAS will require some form of supplement. This supplement should alleviate the problem of radical maneuvers by providing the crew with critical time-based information and the ability to resolve potential conflicts without drastic maneuvers off their published flight path. When a potential conflict is presented to the crew in a timely manner, they can resolve it by performing the necessary deviations to the flight path that do not compromise flight safety or integrity.

Simulation Environment

The goal of the NASA human-in-the-loop simulation was to investigate the feasibility and operational benefits of a concept element (CE) under consideration as part of the DAG-TM program: *CE 5 En Route Free Maneuvering*. The work was completed as part of the Advanced Air Transportation Technologies (AATT) project under NASA's Airspace Systems program. The main purpose of *En Route Free Maneuvering*, is to reduce excessive, en route trajectory deviations that result from separation assurance and traffic flow management (TFM) conformance by distributing the responsibility for separation assurance. An additional benefit of distributing responsibility is that increases in capacity can be realized without placing an added burdening on the ATSP (Raytheon ATMSDI Team. 2004).

The simulation environment was distributed between NASA Ames Research Center and NASA Langley Research Center using the Aeronautical Datalink and Radar Simulator (ADRS) processor to link the facilities. The ADRS functions as the communication management and data distribution hub (Prevot, Palmer, Smith, Callantine. 2002). The DAG-TM airspace was a modified portion of the airspace in and around Fort Worth Air Route Traffic Control Center (ARTCC) (ZFW) and Dallas/Fort Worth TRACON. Participants consisted of seven controllers and 20 licensed pilots. All pilots were air-transport rated and had glass cockpit experience. All of the controllers and 10 of the pilots were located at NASA Ames.

It was important to test this concept in a mixed-equipage environment where some of the participating traffic would be under ATC control (labeled IFR), and others as free flight (labeled AFR for autonomous flight rules). All the AFR pilot-stations in the simulation were equipped with a CSD with datalink capability. Based on the distributed control model, if an AFR aircraft was in conflict with an IFR aircraft, the IFR aircraft always maintained right-of-way, *except* when the ATSP assumed verbal responsibility for conflict resolution. Because of the strategic nature of most conflicts either one or both of the two aircraft may maneuver for resolution, although only one will be ‘burdened”. This procedure allows maximum flexibility but assures that only one aircraft has the *responsibility* for resolving the conflict.

It was expected that all aircraft would remain on their broadcast (assigned) flight path during the simulation, only deviating if commanded by an ATSP. In addition, all AFR aircraft were required to implement only those flight plan changes that would not conflict with the broadcast intent of any other aircraft, IFR or ARF, well beyond the prescribed four minute to LOS window. All burdened AFR aircraft were expected to resolve any predicted high level conflict – notification of a less than four minutes to LOS alert - at least two minutes prior to LOS.

To assist the pilots with making route changes to resolve conflicts, their flight path can be viewed and easily manipulated on the CSD display. Future position over time can be shown with pulse predictor’s running along planned paths of travel. Pilots can also display traffic in a 3D perspective view (Johnson, Battiste, Granada, Johnson, Dao, Wong, Tang. 2005). The route analysis tool (RAT) allows the flight crew to develop, evaluate, and implement potential flight plan changes (Granada, Dao, Wong, Johnson, Battiste. 2005).

Conflict Resolution with CSD Tools

In the DAG-TM studies, an aircraft operating as AFR in the en-route airspace is allowed to free maneuver. This involves the pilot generating user-preferred trajectory changes and instructing the aircraft’s Flight Management System (FMS) to initiate the trajectory. On-board automation broadcasts the modified trajectory using Automatic Dependence Surveillance-Broadcast (ADS-B) to the ATSP and other aircraft. The flight crew has the responsibility to ensure that trajectory changes do not generate near-term conflicts (less than four minutes to loss of separation) with other aircraft. The CD&R provides predicted conflict alerts that require the flight crew to respond accordingly, either taking evasive action or allowing the intruder aircraft to maneuver depending on which aircraft is burdened to resolve (Canton, Refai, Johnson, Battiste. 2005). In contrast to the centralized air traffic management rules that govern IFR operations, in a free flight environment the rules are based on a distributed-control model that references and resolves all potential aircraft conflicts by determining right-of-way, ensuring the pilots participation in the decision-loop in a timely manner.

Normally an aircraft operating in IFR conditions is under the control of an ATSP at all times, with the ATSP retaining separation responsibility. In this study, IFR flights were managed through voice and datalink clearances provided by the ATSP with separation responsibility being transferred to the AFR aircraft.

One critical component of any system where control and responsibility is distributed is a set of operating rules that govern the activities of all participants. However, as we move from centralized to distributed roles and responsibilities, changes will be needed to govern and guide interaction between air and ground operators. The distributed control required for a free flight environment requires, among other things, the successful implementation of new ‘rules-of-the-road’ (ROR) to accompany the new information provided by a CSD and its suite of automation tools.

As participants in a distributed-control environment, the role of the pilot changes. In addition to maintaining responsibility for flying the aircraft, the pilot must now make decisions about the flight plan. For example, pilots will have to ask themselves questions such as “Is my route the most efficient for my aircraft? Is it conflict-free? Is it the best route for meeting the assigned required-time-of-arrival (RTA)?” To aid the pilot, many of these tasks have been automated with the integration of the CSD into the cockpit. Since this new environment is no longer one of “see and avoid” but one of complex, articulated flight paths, the pilot can no longer apply simple flight rules to avoid a conflict situation. Formerly, conflict avoidance was based on the location of the aircraft when the conflict was detected, *not* the location of the aircraft when the conflict may occur. So, a new set of rules, based on the following guidelines were written:

- remove any ambiguity about who is responsible for conflict resolution.
- accommodate more complex route geometries.
- reduce the likelihood of conflicts occurring and encourage a more

organized environment (e.g. pilots abiding by the altitude rule so they reduce the likelihood of being burdened in a conflict).

- if a conflict occurs, allow for a more efficient decision strategy.
- assign responsibility *and* create accountability for conflict resolution.

The CSD rules-of-the-road are listed below and are referenced by the computer automation in the order presented:

- IFR aircraft have the right-of-way when in conflict with AFR aircraft, except when the ATSP has assumed verbal responsibility.

- **Aircraft on a flight plan *always* have the right-of-way** when in conflict with an aircraft that is off of its flight plan, "on a vector". An aircraft is on a vector when its broadcast flight plan does not include a destination airport
- **Altitude Rule:** Aircraft have the right-of-way when:

  A - Traveling EAST (based on the magnetic compass of 0 - 179 deg) and flying at an ODD altitude level.

  B - Traveling WEST (based on the magnetic compass of 180 -359 degrees) and flying at an EVEN altitude level.

  An AFR aircraft is burdened if not flying correct direction for altitude. This rule provides natural separation between level east and west bound flights, reducing the possibility of fast closing head-on conflicts between AFR flights.
- **Left/Right Rule (when conflict angle is > 20 degrees):** Aircraft on the right at the point of conflict has the right-of-way during an encounter between two aircraft when both are level, on ascent, or descent paths.
- **Level Flight Rule:** Aircraft in level flight have the right-of-way over a climbing or descending aircraft (regardless of heading).
- **Descend/Climb Rule:** Descending aircraft have the right-of-way over climbing aircraft. The decision to provide priority to the descending aircraft was to aid flight crews arriving into busy terminals to meet ATM arrival constraints.
- **Overtake Rule**: When the intercept angle between two conflicting aircraft is less than 20 degrees (in other words, they are on the "same" path), the lead aircraft has the right-of-way.

Note: *When none of the rules above apply to the conflict, Ownship assumes responsibility for resolving the conflict. The above rules should cover all possible conflicts, this failsafe rule was added to provide an additional layer of safety. A final failsafe in the system is TCAS.*

Safety of flight takes precedence over all rules.

In addition to the flight rules, the pilots were also assigned specific "roles and responsibilities". Although not embedded in the automation, it was expected they would be followed during the simulation:

- Aircraft ***must*** maintain a minimum separation of 5NM and 1000ft vertical separation from *all* aircraft.
- AFR aircraft ***must*** resolve all conflicts for which they are responsible at least two minutes before LOS. If unable to do so, they were asked to contact the ATSP for assistance.
- AFR aircraft may not create flight plan changes that cause a LOS of less than four minutes.
- The ATSP may verbally assume responsibility for separating an IFR aircraft from an AFR aircraft.
- If the **ATSP** creates a predicted LOS that is within four minutes, the ATSP shall verbally assume responsibility for separating the IFR aircraft from the AFR aircraft.

Although the rules were consistent with those normally used by controllers, they were new to the pilots who informally reported them as cumbersome and difficult to remember. Also, some of the conflict resolution logic may have been counter-intuitive to the pilots who are accustomed to assessing conflict geometry at the time of conflict. Since the rules needed to cover all possible conflict situations that might occur during all phases of flight, reducing the rule set was not an option. And, because it was not the intention of the system designers to turn the pilots into controllers, automating the rule set seemed reasonable.

The automation of the rules relieved the flight crews from the mental and temporal demands of having to assess the right-of-way during a traffic conflict. The knowledge-based system assessed the conflict situation and determined which aircraft was responsible for resolving the conflict. The outcome was a *burden settlement* advisory (Figure 1) to each aircraft involved in the conflict. The burden settlement informs the pilot which aircraft has been burdened with the responsibility to modify their flight trajectory in order to maintain spatial separation (5 NM lateral and 1000 ft vertical) from the conflicting traffic. Each settlement is accompanied by a short phrase, displayed on the CSD, citing the particular rule-of-the-road leading to the settlement advisory. As stated earlier, with the burdening responsibility clearly assigned, it was expected that the crew of the burdened aircraft would take immediate action to resolve the conflict. Pilots were advised not to wait until the LOS window was down to the critical two-minute warning to resolve the conflict for fear that the un-burdened aircraft may feel obligated to resolve the conflict due to the short time to LOS and both resolve simultaneously towards each other. Also, waiting to resolve the conflict could result in a less-optimal solution, increasing the probability that the aircraft, with now limited solutions could resolve by creating a new conflict with another aircraft. If conflicts occurred between multiple aircraft, the

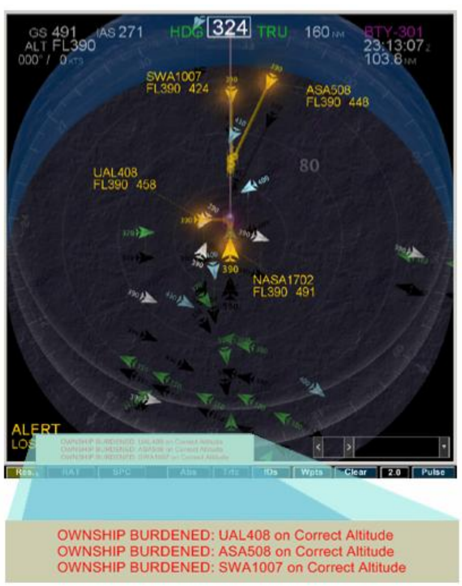


Figure 1. *To remove any ambiguity that may occur when an aircraft is in conflict, the rules clearly assign burdening* responsibility.

burdening logic considered the most eminent conflicting pair by time to LOS, and assigned burdening to one. Once the initial conflict was resolved, the next conflict was considered and a new burdening assignment given. This process continued until all of the potential conflicts were resolved in a timely manner.

Figure 2 shows two AFR aircraft on conflicting routes, one aircraft with an articulated flight path. At the *detection* of the conflict, Ownship is on the right of the conflicting aircraft but at the point of expected LOS, Ownship is on the left and therefore burdened to resolve the conflict. Assessing each conflict situation based on flight path intent and burdening logic does not change over time, or as a function of when the conflict is detected, the rules correctly determine the burdened aircraft at the point of conflict.

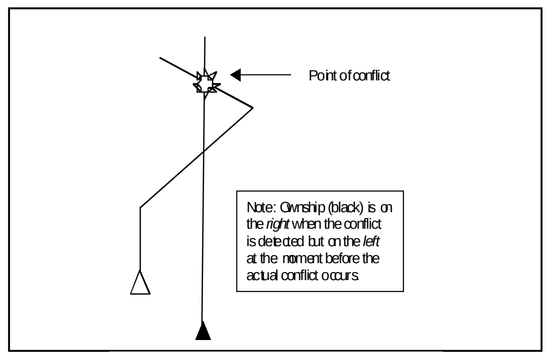


Figure 2. *Example conflict*

Conclusion

After the completion of the simulation, the Ames and Langley pilots participated in separate debrief sessions. The ten Ames pilots were asked in a post-simulation questionnaire, if they thought the rules-of-the-road were "clear and easy to understand", ninety percent (9/10) of the pilots responded 'yes'. When asked if they ever had to mentally reference the rules-of-the-road, forty percent (4/10) responded that they had. Noted observations cited the pilots trying to figure out who would be burdened in a potential conflict before the automation responded with a burdening statement.

The pilots were also asked if the rules were adequate for the mixed equipage (IFR and ARF) environment they were presented, seventy percent (7/10) felt the rules were adequate. The pilots also appreciated that it was not necessary for them to remember the rules; when the rules were needed, the automation supplied them. Our data indicates that the crews were able to resolve all conflicts before any loss of separation events were recorded.

Although the CD&R tools allow ample time to resolve conflicts before the four-minute window, it could be problematic if multiple aircraft are maneuvering under the four-minute window or maneuvering and creating conflicts with less than four minutes to LOS. This circumstance could possibly lead to decisions that create additional conflicts of less than four minutes and therefore resolutions that are less than optimal. To resolve this, perhaps rule-based cooperative strategies for resolution of near-term conflict (those under four minutes to LOS) should be explored. It also may be the case that the four-minute window should be expanded to six minutes. Further research is needed to determine which of these solutions is appropriate.

It is also a plus for this application that the burdening solution is unique to the burdened aircraft; only one aircraft needs to respond with a flight change unlike TCAS which requires both aircraft to respond. This reduces the possibility of both aircraft responding in a manner that jeopardizes the safety of flight. It should be noted that there were no case in which both pilots of conflicting aircraft acted to resolve the conflict, and there were no instances of competing maneuvers. These performance results, in a simulated free flight environment, suggest that the aforementioned rules-of-the-road can adequately support self-separation. Maintaining separation requires earlier responses than the current-day collision avoidance tools, and therefore a system such as the one described herein may be necessary to support free flight.

References

Bone, R. Domino, D., Helleberg, J., Johnson, N. (2003). Cockpit display of traffic information (CDTI) enhanced flight rules (CEFR) simulations 1-4. (MITRE Technical Report). McLean, Virginia : MITRE Corporation

Canton, R., Refai, M., Johnson, W. W., Battiste, V. (2005). Development and integration of human-centered conflict display and resolution tools for airborne autonomous operations. Submission to the 2005 International Symposium on Aviation Psychology, Oklahoma City, OK

Granada, S., Dao, A., Wong, D., Johnson, W.W., Battiste, V. (2005). Development and integration of a human-centered volumetric cockpit situational display for distributed air-ground operations. Submission to the 2005 International Symposium on Aviation Psychology, Oklahoma City, OK.

Johnson, W.W., Battiste, V., Granada, S., Johnson, N., Dao, A., Wong, D., Tang, A. (2005). A simulation evaluation of a human-centered approach to flight deck procedures and automation for en-route free maneuvering. Submission to the 2005 International Symposium on Aviation Psychology, Oklahoma City, OK.

Prevot, T., Palmer, E., Smith, N., Callantine, T. (2002). A Multi-Fidelity Simulation Environment for Human-in-the-Loop Studies of Distributed Air Ground Traffic Management. American Institute of Aeronautics and Astronautics Modeling and Simulation Conference and Exhibit, Monterey, CA.

Raytheon ATMSDI Team. (2004). Air traffic management system development and integration (ATMSDI) NASA Joint DAG-TM/CE 5 simulation test plan. Draft document, NASA Ames Research Center, Mountain View, CA.

Radio Technical Commission for Aeronautics (RTCA). (1995b). Free flight implementation (RTCA Task-Force 3 Report). Washington, D.C.: RTCA Inc.

http://www.aerowinx.de/html/tcas.html

http://www.fact-index.com/v/vi/visual_flight_rules.html

http://www.atsb.gov.au/aviation/occurs

OBJECTIVE PILOT PERFORMANCE MEASUREMENT: A LITERATURE REVIEW AND TAXONOMY OF METRICS

Nicholas R. Johnson and **Esa M. Rantanen**
University of Illinois at Urbana-Champaign, Aviation Human Factors Division
Savoy, Illinois, USA

This paper will present a review of development and evaluation of objective pilot performance measures. A basic taxonomy of measure types is also presented where objective measures are classified distinguishing between direct measures, derivative metrics based on these, and indirect measures of unobservable constructs and phenomena. The recent development of novel objective measures and their ability to predict and classify pilot performance will also be reported.

Introduction

Traditional means of evaluating pilot performance have typically involved the subjective evaluation of an instructor or highly experienced pilot. Performance assessment based on subjective evaluations has the advantages of relative ease of implementation, high face validity, and simplicity in providing specific feedback to the examinee pilot. However, subjective evaluations of performance are prone to the problems of inter and intra-rater reliability. Additional factors such as group performance levels and raters' temporal variance of performance standards may also influence the reliability of subjective evaluations of pilot performance. Objective measures based on flight data recordings have the potential to alleviate these concerns and their use can provide an alternative or complimentary approach to pilot performance measurement in training and research environments. The utility of objective performance measures includes use in transfer of training studies, validation of training methods, development of automated or adaptive training methods, logging of solo flights for subsequent evaluation, and standardized check rides. In addition, by alleviating the time constraints and information overload often associated with direct observation, automated data collection can enhance and expand traditional proficiency evaluation methods by an instructor pilot. Furthermore, quantitative performance data can be utilized in research and subjected to various statistical analyses to reveal underlying, covert patterns in pilots' performance.

Review of Metrics

Although our literature review was aimed to be exhaustive, much of the past military research may not have been published and hence not been available for review. Other reviews of objective pilot performance measures include Mixon and Moroney (1982), who listed 189 articles on objective pilot performance measurement, broken down into fixed/rotary wing aircraft and simulator/field studies. No attempt was made to review or critique the articles, but the number of subjects, equipment, scenarios and measures were listed. Gawron's (2000) handbook of performance measures includes objective measures of however, the major focus of the review is on workload measures.

Basic Measures

Ideally, objective measures should be temporally invariant (repeatable), criteria based, insensitive to group performance, transferable between simulator and aircraft, interpretable, and for training purposes, be immediately available for feedback or monitoring. Basic measures such as Root Mean Square Error (RMSE) and Standard Deviation (SD) satisfy most of these requirements. However they do not contain information about the direction of the deviations or the frequency of these deviations from the mean. Consequently, identical numeric values for RMSE (or SD) can result from quite distinct performance. Hubbard (1987) noted that by using the mean and the SD together instead of the RMSE, a more complete picture of performance could be obtained, albeit at a cost of having to simultaneously interpret two measures instead of just one.

Measures based on amplitudes of flight parameter data can provide additional information on pilot performance when combined with tolerance values. Number of deviations (ND) and time spent outside parameter tolerances (TD) are an example of such measures. The ND is a measure that tallies the occurrences of the aircraft straying outside predetermined tolerances (Reynolds, Purvis, & Marshak, 1990). This is essentially a measure of velocity error in tracking and it complements the RMSE, which contains the error magnitude information. A low number typically indicates good performance. A low value, however, can also be obtained if the pilot makes few aberrations outside the tolerances but stays there for a substantial proportion of the flight segment of flight. The ND measure must hence be considered together with the total time spent outside tolerance in a given segment. The TD measure provides an indication of tracking performance beyond the RMSE and ND measures. TD is computed simply by summing the time the pilot spends outside of a given tolerance and divided by the total time in the segment (i.e., percent time outside tolerance). A small number indicates good performance. Sirevaag, Kramer, Wickens, Reisweber, Strayer and Grenell (1993) took aircraft control measures from a helicopter simulator in a study investigating the effects of verbal and digital communication loads on pilot performance. The measurement of time above an altitude criterion produced significant differences between experimental task conditions.

Rantanen and Talleur (2001) developed a metric labeled mean time to exceed tolerance (MTE). The MTE is computed from the rate of change between successive data points and the aircraft's position relative to a given tolerance. Based on this information, the measure extrapolates the time the aircraft will remain within the tolerance region. In subsequent analysis, the MTE from tracking the localizer on an instrument landing system (ILS) approach showed a significant difference between pilots who passed an instrument proficiency check flight and those who failed, by flight instructor evaluation (Rantanen & Talleur, 2001).

Novel Measures

The above measures are relatively simple, have a high degree of face validity and are easily interpreted. Attempts have also been made to develop more novel measures, based on somewhat more complicated constructs or data analysis.

De Maio, Bell, & Brunderman (1985) defined a critical control input as a pilot control input that changed or led to a change from positive vertical acceleration to negative vertical acceleration (or other flight parameter) or vice versa. Conversely, a non-critical control input did not cause the vertical acceleration to change from positive to negative or vice versa. The authors hypothesized that "efficient" control would be characterized by a relatively large proportion of critical control inputs indicating that pilots were canceling small errors in altitude frequently. A measure of "smoothness," was subsequently defined as the proportion of critical control inputs from the total number of inputs (critical + non-critical). The critical error rate is the horizontal distance traveled from critical control input to vertical acceleration sign change divided by the time from critical control input to vertical acceleration. This metric was designed to measure the effectiveness of a critical control input; low values for critical error rate would indicate a slow accumulation of error following the pilot control input. De Maio et al. (1985) found that that smoothness and critical error rate were affected by flight task difficulty (straight vs. turning flight, both at constant altitude).

The n^{th} moment of a series of data is the summation of individual series values raised to the n^{th} power and then divided by the number of sample points. Thus, the first moment is simply the average of a series of data. Average values have been commonly used as measures of pilot performance; for example, Hills and Eddowes (1974), McDowell (1978) and Sirevaag et al. (1993). However their use may be limited in certain circumstances given the way averages can mask important patterns and deviations in performance. The use of higher order moments appears to have been limited to McDowell (1978), where the aileron second moment showed differences between pilot experience groups in the study.

Gaidai and Mel'nikov (1985) developed a measure based on an integral equation to evaluate pilot performance in a landing task. The measure took the weighted sum of normalized deviations from criterion values over several flight parameters. The explicit form of the equation can also be found in Gawron (2000).

Frequency analysis (Bloomfield, 1976; Gottman, 1981) has also been identified as a useful tool to aid in performance measurement (Semple, Cotton & Sullivan, 1981; Benton, Cooriveau & Koonce, 1993). However, actual implementations of such frequency-based measures have been limited. Hills and Eddowes (1974) and Vreuls, Wooldridge, Obermayer, Johnson, Normal & Goldstein (1976) used measures based on a manual tracking approach. Given a known disturbance function that was applied to the simulator aircraft, the researchers were able to use control inputs and derive Bode plots of pilot performance. From this, measures such as cross-over power and high- and low-frequency gains were generated. Hills and Eddowes used these measures as part of a battery of over 2000 measures that attempted to classify pilot experience groups . Vreuls et al. (1976) performed a similar analysis, however the exact nature of the frequency-based measures that were included is not clear. By contrast, McDowell (1978) did not use a manual tracking approach and instead used several measures to quantify pilots' control input power spectra. Several "digital filter" type measures were developed that estimated the relative power spectra below various frequency cut-off points. The 0.125Hz cut-off filter measure from the aileron control inputs produced the greatest separation between pilot experience groups of these filter metrics. In this case the more skilled pilots had their power spectra shifted towards higher frequencies.

Following a similar approach to McDowell (1978), Johnson, Rantanen and Talleur (2004) developed a number of frequency analysis based measures. Using data collected from instrument proficiency check flights in an aircraft and two types of flight simulators, Johnson et al. derived seven distinct measures of performance across nine different flight parameters. The authors found that many of the measures were sensitive to differences in pilot performance as judged by an instructor pilot (IP). Specifically, measures of the mean and standard deviation of the magnitude of components in the frequency distribution of a flight parameter's time series data were found to be sensitive to pilot performance. Using the same data, Rantanen, Johnson and Talleur (2004) found such differences in pilot performance were most clearly seen during localizer and glide-slope tracking on an ILS approach. In addition, low-pass filter measures were implemented in a similar manner to McDowell (1978). While these measures were not, in general, as effective at separating pilot performance groups (as judged by an IP), their sensitivity may have been limited by not setting the cut-off frequency optimally.

Combining Metrics

Further objective measure development has included techniques based on combining individual flight parameter measures into an index of pilot performance. Knoop and Welde (1973) used a summation of absolute values of flight parameter deviations from criterion values at four chosen points in a lazy-8 flying maneuver performed in a T-37 military training aircraft. The flight parameters used in this index included airspeed, altitude, heading, pitch, roll and pitch, roll and yaw rates. The index was compared to subjective evaluations of an instructor pilot (IP) where it was found that the index accounted for 67% of the variance of the IP's ratings. In a parallel study, Knoop (1973) also introduced the idea of a linear combination of Boolean measures based on pilot performance within flight parameter tolerance ranges. That is, a 1 or 0 would be scored by the pilot depending on whether they flew the aircraft within the acceptable range of flight parameter value. Problems of intra- and inter-rater reliability made meaningful comparison of this index with subjective evaluations difficult. Childs (1979) also used an index based on categorical values in assessment of helicopter pilot performance. Altitude, airspeed and heading performance were evaluated and a score from 1 to 6 was assigned in each flight segment based on whether performance was within one of three tolerance bands. These segment scores were then combined to form an overall flight score from 1 to 6. Results indicated this measure was sensitive to training time, but no other validation was attempted.

Connelly, Bourne, Loental, Migliaccio, Burchick and Knoop (1974) performed a study that was concerned with developing candidate measures for pilot performance evaluation in a T-37 aircraft. They studied lazy-8, approach and landing, barrel roll, split-S, and cloverleaf maneuvers, specified the flight parameters and control inputs to be recorded and developed measures for performance evaluation. These measures were formulated in terms of continuous differences from a reference trajectory, where this trajectory could be empirically derived, and tolerance values were based on either external criterion or SDs from empirical data. Linear combinations of weighted errors (c.f., Knoop & Welde, 1973) and vector combination of error terms (allowing simultaneous comparison of all error terms) were discussed but because no data was collected, these combinations could not be evaluated.

Bortolussi and Vidulich (1991) developed a figure of merit (FOM) of pilot performance from six primary flight variables (control inputs, altitude, airspeed and heading). The authors studied both a total FOM (derived from standard deviations of the six variables and the altitude, airspeed and heading means) and specific flight parameter FOMs. For example, an altitude FOM was derived from altitude mean and SD and from the SD of elevator inputs. The FOMs were produced by a weighted linear combination of their component primary variables. The authors found the weighting coefficients by using an analytical hierarchy process. The FOMs were evaluated by comparison between two flight scenarios of differing difficulty. While the total FOM was not sensitive to scenario difficulty, the altitude and airspeed FOMs did differ between conditions.

It is evident that complex tasks—such as flying—involving multiple dimensions (exemplified by flight parameters in our discussion) can yield a vast number of measures. In addition to the efforts described above to combine measures, several attempts have been made to statistically reduce the number of measures using discriminant analysis (Hills & Eddowes, 1974; Vreuls et al, 1975; Kelly, Wooldridge, Hennessy and Reed (1979). For example, Hills and Eddowes' (1974) study yielded a total of 2436 measures per subject. The authors attempted to distinguish three pilot experience groups based on the objective measures derived from the flight tasks. One-way ANOVAs were used to determine the ability of each measure to independently predict pilot group membership. Only a little over 17 % (420) of the variables were found to be statistically significant in separating groups. Standard deviation measures produced the highest proportion of significant variables (32%), followed by frequency analysis measures (20%), means (18%) and correlations (11%). A linear discriminant function that was derived from the results of the first experiment was used to classify performance in the second experiment with new subjects. The classification process was statistically successful. However, the discriminant function misclassified 33% of subjects, leading the authors to question the practicality of using discriminant functions to diagnose performance.

Vreuls et al. (1976) also sought to limit the number of performance measures and utilize those that could discriminate between early and late training in an automated IFR training simulator. Basic measures and measures derived from frequency analysis of standard flight parameters were used to generate a discriminant function. The discriminant function contained 9 derived measures on average, including several control input variables. Using these discriminant functions in automated feedback training scenarios reduced training time to set criteria by 34–40% compared to the original method that was not based on a discriminant function. The mixed results of Hills and Eddowes (1974) and Vreuls et al. (1976) highlight the difficulty in reducing a large number of performance measures into a manageable set that can be used to reliably predict skill level or measure performance.

Taxonomies of Measures

In addition to statistical techniques, some kind of classification system should be considered to help manage the large number of metrics. A thorough review of past and current research efforts and organization of the findings in a manner that facilitates the use of existing knowledge is critical for future evolvement of pilot performance measurement. On one hand, this helps to avoid 'reinventing the wheel.' On the other hand, periodic literature reviews provide for a foundation for future research efforts by defining a 'toolbox' of measures that would predict pilots' success in their task and the impact of changing training protocols, procedures, and new technology on the system as a whole.

The basic structure of the taxonomy of measures proposed here is classification by flight parameters and distinguishing between *direct* measures, *derivative* metrics based on these, and *indirect* measures. Direct measures are momentary values of flight parameters, for example, altitude or heading. Derivative metrics are based on these, for example, mean and standard deviation of altitude values. Indirect measures are those that cannot be measured directly but must be inferred from derived measures, for example, *pilot performance* based on standard deviation of altitude in level flight.

Table 1 depicts the directly measured variables found in the literature, the frequency of their encounters, the percentage of all parameters, and a cumulative percentage. Altitude, airspeed, roll, control inputs, heading and pitch were the most frequently measured variables, together accounting for over 65% of all parameters measured.

Table 1. *Frequencies of flight parameters.*

Parameter	Freq.	%	Cum. %
Altitude	21	12.88	12.88
Airspeed	19	11.66	24.54
Roll	17	10.43	34.97
Control Inputs	17	10.43	45.40
Heading	16	9.82	55.21
Pitch	16	9.82	65.03
Vertical Speed	11	6.75	71.78
VOR Tracking*	8	4.91	76.69
Yaw	5	3.07	79.75
Turn Rate	5	3.07	82.82
Glide Slope Tracking	5	3.07	85.89
Flaps	4	2.45	88.34
Trim	4	2.45	90.80
Speed Brakes	3	1.84	92.64
Sideslip	3	1.84	94.48
Landing Gear	3	1.84	96.32
Acceleration	3	1.84	98.16
Position	2	1.23	99.39
NDB tracking**	1	0.61	100.00

*VOR = Very High Frequency Omnidirectional Range
** NDB = Non-directional Beacon

There are two main issues to consider when interpreting these results. First, the ease and practicality of making measurements of any particular parameter clearly plays a role in their ranking in Table 1. Second, the relevance of the parameters depends heavily on the particular flight maneuver to be evaluated. For example, altitude measurements may yield little useful information if the pilot is climbing or descending (c.f., Rantanen, Johnson, & Talleur, 2004).

The second major class, derivative measures, may be further divided into several subclasses according to the particular (mainly statistical) techniques used to reduce the often massive amounts of data into something manageable and interpretable. These derivative metrics are depicted in Table 2, again ranked by the frequency they were encountered in the literature. Not surprisingly, RMSE ranked first, followed by SD, maximum and minimum values and mean.

Table 2. *Derivative measures used in the literature.*

Derivative Metric	Freq.	%	Cum. %
RMSE	16	21.92	21.92
Std. Dev.	8	10.96	32.88
Max/min	8	10.96	43.84
Mean	6	8.22	52.05
Frequency Analyses	5	6.85	58.90
Range	5	6.85	65.75
Deviation from criterion	4	5.48	71.23
Time on target	4	5.48	76.71
Mean absolute error	3	4.11	80.82
Autocorrelation	3	4.11	84.93
Time outside tolerance	3	4.11	89.04
Median	2	2.74	91.78
ND	2	2.74	94.52
Boolean	1	1.37	95.89
Correlation	1	1.37	97.26
Moments	1	1.37	98.63
MTE	1	1.37	100.00

Finally, the third main class of measures, indirect measures, only had one subcategory: pilot performance. What is noteworthy, however, is that very little was found in our literature review that would link direct measures to the measures of real interest, that is, performance, via a valid or even plausible theoretical construct. As availability of data from flight data recorders and data outputs from ground-based trainers is not a problem, and as there exists many established techniques to process and reduce these data to metrics (c.f., Table 2), the lack of theoretical foundation for measurement of pilot performance is conspicuous.

Another measure classification scheme to help in data reduction and interpretation is based on task analysis of piloting an airplane. The navigational goals of a pilot (e.g., a given heading, altitude, or track over ground) are hierarchical (Wickens, 2003). Furthermore, the control order changes across the hierarchy, and, given that humans have increasing difficulty controlling higher order systems, it is important to recognize what is the appropriate parameter to control in a given task or situation. For example, aircraft altitude control depends on the zero-order control of elevator angle, which results in the first order control of pitch angle, which in turn affects the second order control of vertical speed of the aircraft, which finally determines the aircraft's altitude, which can be seen as a third order control task. Obviously, other controls are coupled with this task, for example, engine thrust (zero-order) and airspeed (first order), further complicating the pilots' task. However, such hierarchy offers a promising framework for the choice, analysis, and interpretation of objective metrics available from different maneuvers.

Discussion

This review has highlighted both standard objective measures and attempts to develop novel diagnostic measures of pilot performance. Despite a relatively long history and numerous and varied approaches to development of objective pilot performance measures, successes of measure validation for all but the most basic metrics have been limited. Efforts to corroborate the effectiveness of objective measures in describing pilot performance have focused on the measures' sensitivity to training, correlations with subjective evaluations, or performance in cross-validation studies.

While some of the mathematical techniques described in this paper offer the potential to uncover detail or patterns in pilot performance that may not be perceptible or quantifiable to a human observer, the task of flying an airplane is such a multi-faceted one that simply looking at a single flight parameter may not yield much diagnostic information. Instead, it appears that the greatest potential for diagnostic objective indices lies in the formation of measures combined from various related direct measures. Such combinations should be based on a detailed analysis of the flying task involved and utilize the natural linking of flight parameters though the hierarchical structure of pilot goals and control order. Even if combinations of objective measures fail to produce the performance sensitivity and diagnosticity required for research and training purposes, they can still be used to assist pilot performance evaluation.

References

Benton, C.J., Corriveau, P., & Koonce, J.M. (1993). *Concept development and design of a semi-automated flight evaluation system (SAFES*). [AL/HR-TR-1993-0124]. Brooks AFB, TX: Armstrong Lab., Human Resources Directorate.

Bloomfield, P. (1976). *Fourier analysis of time series: An introduction*. New York: Wiley.

Bortolussi, M.R. & Vidulich, M.A. (1991). An evaluation of strategic behaviors in a high fidelity simulated flight task. Comparing primary performance to a figure of merit. *Proceedings of the 6th ISAP*, 2, 1101-1106.

Childs, J.M. (1979). The development of objective inflight performance assessment procedures. *Proc. 23rd HFES Mtg.* Santa Monica, CA: HFES.

Connelly, E. M., Bourne, F. J., Loental, D. G., Migliaccio, J. S., Burchick, D. A. & Knoop, P. A. (1974). *Candidate T-37 pilot performance measures for five contact maneuvers* [AFHRL-TR-74-88]. WPAFB, OH: AFHRL

De Maio, J., Bell, H. H., & Brunderman, J. (1985). *Pilot-oriented performance measurement*. [AFHRL-TP-85-18]. Brooks AFB, TX: AFHRL.

Gaidai, B.V. & Mel'nikov, E.V. (1985). Choosing an objective criterion for piloting performance in research on pilot training on aircraft and simulators. *Cybernetics and Computing Technology*, 3, 162-169.

Gawron, V. J. (2000). *Human performance measures handbook*. Mahwah, NJ: Erlbaum.

Gottman, J. M. (1981). *Time series analysis: A comprehensive introduction to social scientists*. Cambridge, UK: Cambridge University Press.

Hills, J. W., & Eddowes, E. E. (1974). *Further development of automated GAT-1 performance measures*. [Rep 73-72]. Brooks AFB, TX: AFHRL.

Hubbard, D.C. (1987), Inadequacy of root mean square error as a performance measure. *Proceedings of the 4 th ISAP*. Columbus, OH: OSU.

Johnson, N. R., Rantanen, E. M., & Talleur, D. A. (2004). Time series based objective pilot performance measures. *International Journal of Applied Aviation Studies* (IJAAS), 4(1), 13-29.

Kelly, J. K., Wooldridge, A. L., Hennessy, R. T. & Reed, J. C. (1979) Air combat maneuvering performance measurement. *Proceedings of the Human Factors Society 23rd Annual Meeting*. Santa Monica, CA: HFES.

Knoop, P. A. (1973). Advanced instructional provisions and automated performance measurement. *Human Factors*, 15 (6), 583-597.

Knoop, P. A. & Welde, W. L. (1973). *Automated pilot performance assessment in the T-37: a feasibility study* [AFHRL-72-6]. Wright Patterson AFB, OH: AFHRL.

McDowell, E. D. (1978). *The development and evaluation of objective frequency domain based pilot performance measures in ASUPT*. Bollings AFB, DC: Air Force Office of Scientific Research.

Mixon, T. R., & Moroney, W. F. (1982). An annotated bibliography of objective pilot performance measurements [NAVTRAEQIPCEN-IH-330]. Orlando, FL: Naval Training Equipment Center.

Rantanen, E.M., Johnson, N.R., & Talleur, D.A. (2004). *The effectiveness of a personal computer aviation training device, a flight training device, and an airplane in conducting instrument proficiency checks, Vol. 2: Objective pilot performance measures* (AHFD-04-16/FAA-04-6). Savoy, IL: AHFD

Rantanen, E. M., & Talleur, D. A. (2001). Measurement of pilot performance during instrument flight using flight data recorders. *International Journal of Aviation Research and Development*, *1*(2), 89-102.

Reynolds, M. C., Purvis, B. D., & Marshak, W P. (1990). A demonstration/evaluation of B-1B flight director computer control laws: A pilot performance study. *Proceedings of the IEEE National Aerospace and Electronics Conference* (pp. 490-494). Piscataway, NJ: IEEE

Semple, C. A., Cotton, J. C. & Sullivan, D. J. (1981). Aircrew training devices: Instructional support features [AFHRL-TR-80-58]. Brooks AFB, TX: Air Force Human Resources Laboratory.

Sirevaag, E. J., Kramer, A. F., Wickens, C. D., Reisweber, M., Strayer, D. L., & Grenell, J. F. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics*, 36(9), 1121-1140.

Vreuls, D., Wooldridge, A. L., Obermayer, R. W., Johnson, R. M., Norman, D. A., & Goldstein, I. (1975). Development and evaluation of trainee performance measures in an automated instrument flight maneuvers trainer. [NAVTRAEQUIPCEN 74-C-0063-1]. Orlando, FL: Human Factors Laboratory, Naval Training Equipment Center.

Wickens, C. D. (2003). Pilot actions and tasks: Selections, execution and control. In P. S. Tsang and M. A. Vidulich (Eds.), *Principles and Practice of Aviation Psychology* (pp. 239-263). Mahwah, NJ: Lawrence Erlbaum.

Wooldridge, L., Obermayer, R.W., Nelson, W.H., Kelly, M.J., Vreuls, D. and Norman, D.A. (1982). *Air combat maneuvering performance measurement state space analysis* [AFHRL-TR-82-15]. Brooks AFB, TX: AFHRL

COCKPIT TECHNOLOGY AND WEATHER RELATED DECISION MAKING: AN INTEGRATIVE REVIEW

Johnson, N.R., Wiegmann, **D.A., Goh, J.** and **Wickens, C.D.**
Aviation Human Factors Division, Institute of Aviation
University of Illinois at Urbana-Champaign
Savoy, IL

This paper provides a synthesis of the empirical studies to date that have investigated weather-related pilot decision making. Of particular interest is how new cockpit technologies such as synthetic vision systems and graphical weather information systems interact with flight experience, risk perception, risk-taking tendencies and self-confidence in affecting pilots' decision to continue flight into adverse weather. A conceptual framework for integrating and interpreting the results of these various studies is also be proposed.

Introduction

Visual flight rules (VFR) flight into instrument meteorological conditions (IMC), or unqualified flight into adverse weather, continues to be a major safety hazard within general aviation (GA). In an analysis of GA accidents between 1990 and 1997, Goh and Wiegmann (2002) found that the fatality rate of VFR into IMC accidents was approximately 80%. This compared to a fatality rate of approximately 19% for other types of GA accidents during the same period. These statistics reflect similar trends found by the National Transportation Safety Board (1989) for United States GA accidents that occurred during the 1970s and mid-1980s, as well as GA accident trends in other countries (e.g., United Kingdom and New Zealand; O'Hare & Smitheram, 1995). In a recent analysis, Knecht, Harris and Shappell (2004) found that the GA fatality rate per million passenger miles was 223 times that of commercial aviation for the period 1990-1998. In addition, the authors' analysis revealed that IMC was implicated in 32% of these GA fatalities. In sum, these findings clearly indicate that VFR flight into IMC continues to be a major safety hazard within general aviation.

Visual flight rules flight into IMC is often characterized by pilots' decisions to continue a flight into adverse weather conditions, despite having been given information or presented with cues that indicate they should do otherwise (NTSB, 1989). One possible explanation for VFR flight into IMC is based on the predictions made by prospect theory (Kahneman & Tversky, 1982). The hypothesis put forward by O'Hare and Smitheram (1995) and O'Hare and Owen (1999), predicts that pilots framing the decision to continue flight into deteriorating weather in terms of losses (e.g. loss of time, money, effort) will be risk-seeking and continue with the flight. Conversely, those pilots who frame the decision in terms of gains (e.g. personal safety) will be risk-averse and divert the flight when faced with deteriorating weather. This hypothesis was supported by analysis of GA accidents in New Zealand over the period 1988 to 2000 (Owen, O'Hare and Wiegmann, 2001). The authors' found that weather-related accidents occurred significantly farther from the departure point than other types of accidents (loss of control, collision with terrain and mechanical failure). Efforts to demonstrate a similar pattern in laboratory studies have been unsuccessful however.

There is a growing body of evidence that errors in situation assessment may lead to pilots' decisions to continue flight into deteriorating weather. Goh and Wiegmann (2001) and Wiegmann, Goh and O'Hare (2002) found that VFR pilots who continued simulated flights into adverse weather generally misinterpreted weather information, overestimating weather parameters. That is, pilots who continued flight had more positive views of cloud ceiling and visibility than those pilots who did not continue the flight. Pilots involved in VFR into IMC accidents also generally have less experience diagnosing and flying in adverse weather (Goh & Wiegmann, 2002). Burian, Orasanu & Hitt (2000) found that pilots in their study who were in the 25th percentile and below in terms of total flight hours were more likely to press on into deteriorating weather than those in the 75th percentile and above. The authors suggested that some pilots, particularly those with less experience, "do not trust what their eyes are telling them and so proceed on blindly"(p. 25). Therefore, at least in some situations, VFR flight into IMC can be viewed as a failure in recognition-primed decision making (RPD; Klein, 1993). Consequently, training and technological inventions that focus on improving pilots' situation awareness (SA) and weather evaluation might improve pilot decision making, thereby reducing accidents due to VFR flight into IMC.

Contrary to the above evidence, however, are findings indicating that some pilots occasionally choose to continue flight into adverse weather even after they have become aware of the hazardous conditions (Burian, Orasanu & Hitt, 2000). Pilots who continue flight into adverse weather tend to be overconfident in their abilities and also underestimate the risks of VFR flight into IMC (O'Hare, 1990). Indeed, results from the Goh and Wiegmann (2001) study partially support this hypothesis in that pilots who chose to continue a simulated flight into adverse weather were more confident in their skills compared to those who chose to divert and generally underestimated the risks of crashing due to the weather.

Some researchers have found that prior exposure to adverse weather improves pilots' situation assessment abilities (Wiggins & O'Hare, 1995; O'Hare, Owen & Wiegmann, 2001) but also reduces their perceptions of risk. In a simulator study involving a 180° turn out of IMC, Goh and Wiegmann (2004) found that pilots who flew the turn in a low-turbulence condition had reduced perceptions of risk compared to pilots who flew in a high-turbulence condition. In subsequent encounters with adverse weather, those pilots with reduced risk perception may be more willing to fly into deteriorating weather or enter into flight conditions that exceed their abilities.

Novacek, Burgess, Heck and Stokes (2001) found that pilots who possessed more extreme risk-taking personalities were also more likely to make riskier/poorer weather-related decisions when using a NEXRAD display than those pilots who were generally risk averse. Collectively, these findings suggest that efforts to improve pilots' weather related decision making should not only address situation awareness and assessment but also the potential impact of such efforts on risk-taking behavior.

Unfortunately, only a few empirical studies to date have been conducted to examine the impact that different types of technology aboard aircraft have on GA pilots' decisions to continue VFR flight into IMC. In one such study, O'Hare, Owen and Wiegmann (2001) investigated GA pilots' use of a global positioning system (GPS) during a cross-country flight in deteriorating weather conditions. The authors found that pilots flying an airplane equipped with a global positioning system (GPS) were more accurate in their position assessments and had a greater confidence in their location than pilots who flew without GPS equipment. In addition, the pilots with GPS were more likely to continue flight into IMC or remain airborne longer than pilots without GPS. The results of the study also showed that pilots who continued flight into IMC had lower estimates of the risks involved, compared to pilots who diverted, corroborating the previous findings of Goh and Wiegmann (2001).

In another study, Beringer and Ball (2003) investigated how variations in the data resolution of an on-board graphical weather information system (GWIS) affected pilots' judgment of weather severity and decisions to continue a simulated cross-country flight. The GWIS used NEXRAD (NEXt-generation-RADar) data to give pilots graphical information on the location and intensity of local area precipitation. The NEXRAD data was presented in 8km, 4km or 2km resolutions. The authors found that pilots with the highest resolution NEXRAD display (2km) spent the most time looking at the GWIS display, delayed their decision to divert the longest and came closest to the thunderstorm cells compared to the other two lower-resolution pilot groups. Based on these results and further data from post-flight static image judgments, the authors suggested that the high-resolution NEXRAD displays are likely to encourage pilots to continue flight while attempting to maneuver around or between the significant weather cells.

These findings suggest that as weather and other navigation displays become more advanced and sophisticated they may shift pilots' decision making processes from that of *strategic* decision making to that of *tactical* decision making. A pilot using such a display strategically may attempt to avoid a hazard altogether, whereas a pilot using a display tactically may attempt to negotiate a path through a weather hazard area such as a broken line of thunderstorms. In general, the distinction between the two types of decision making is that tactical decision making will be reactive to immediate environmental events while strategic decision making will be proactive and include planned avoidance of potentially hazardous events (Latorella and Chamberlain, 2002). Such shifts in decision strategies could have severe negative ramifications for generally less-skilled GA pilots.

A particular advance in cockpit technology that could affect these GA pilots in the near future is the synthetic vision system (SVS). SVS displays provide the pilot with an ego-centric, synthetic realization of terrain and other potential hazards (for example, traffic or towers) in front of the aircraft to better support flight in challenging terrain or low visibility conditions. Typically, the SVS display will include flight path guidance in the form of a highway in the sky (HITS; Alexandra, Wickens and Hardy, 2003; Williams, 2002; Berringer, 2000). It is hoped that synthetic vision technology will help prevent controlled flight into terrain (CFIT) and low-visibility loss of control (LVLOC) GA accidents.

Takallu, Wong and Uenking (2002) examined the use of SVS technology to help counter LVLOC accidents. In their flight simulation study, non-instrument rated GA pilots were required to execute basic flight maneuvers after entering into IMC. The maneuvers (180° turn, straight climb, straight descent, straight and level flight) were performed with either standard instruments or a SVS display (without HITS). The authors found that pilots flying with the SVS generally committed fewer violations of the altitude, heading or airspeed tolerances that were specified prior to the flight maneuvers. The improved performance of pilots while using the SVS display was attributed to enhanced spatial awareness that the display afforded.

To the best of our knowledge, there have been no further studies that investigate VFR pilots' use of SVS displays while encountering deteriorating weather. There is however, the opportunity to examine issues with SVS technology that may be relevant to weather related decision making that have been raised in a number of different studies. Although it has commonly been found that a HITS SVS display supports flight path tracking (e.g., Iani & Wickens, 2004; Prinzel, Comstock, Glabb, Kramer, Arther & Barry, 2004) there is evidence that there may be performance trade-offs. For example, the clutter associated with over-laying traffic information, traditional aircraft instrumentation and the HITS may inhibit traffic detection, in particular traffic that is neither expected or salient (Wickens, Ververs and Faden, 2004). In addition, the compelling nature of the HITS SVS may cause pilots to shift a disproportionate amount of visual attention to the SVS display, or at least make it more difficult to switch attention away from the SVS display to perform a concurrent task.

While not specifically looking at an SVS display, Wickens , Goh, Helleberg, Horrey and Talleur (2003) found that a cockpit display of traffic information (CDTI) affected pilot detection of a "rouge" aircraft that was only visible in the outside world. By drawing pilot attention away from the outside world, the CDTI made it more difficult for pilots to detect the rouge aircraft.

In contrast to the hypothesis that a HITS SVS display may be so compelling as to reduce outside-world scanning performance, is the hypothesis that such a display may lead to an increase in concurrent task performance by alleviating the workload associated with the primary flight control task. This hypothesis is supported by the finding that pilots flying with a HITS SVS display were more sensitive to weather changes presented in a secondary cockpit display than pilots without the HITS (Iani and Wickens, 2004). For integration with the issues raised in this paper, it should be noted that weather-event detection occurred from the in-cockpit display and not from any visual cues in the outside world (pilots were flying in IMC).

The manner in which scanning mediates the relationship between display and performance also needs to be considered. While Wickens et al. (2003) found a coupling between reduced outside world scanning and poorer performance in traffic detection, the relationship is not always consistent. Williams (2002) for example, found that despite the fact that the time pilots spent scanning the outside world decreased with a HITS SVS, their ability to detect outside world traffic was not reduced significantly.

The above studies, while not directly addressing VFR flight into IMC, highlight some general issues with pilot use of synthetic vision systems that are worth considering in a weather related decision making context. If pilots' scanning behavior is altered by the new technology, resulting in less time spent looking outside the cockpit, it is reasonable to expect that pilots' weather situation assessment may become poorer. Also, if the HITS appears to present the pilot with enough flight path information to navigate without reference to the outside world, it may encourage certain pilots to fly into deteriorating weather believing they can use the technology exclusively. On the other hand, if pilot flight control workload is reduced with the HITS SVS display, it is reasonable to expect that weather-related decision making may improve as more mental resources are available to integrate the in-cockpit weather information and the outside world weather cues.

The specific parameters of advanced displays that impact pilot-decision making, however, have yet to be systematically identified. Hence, little is known about how to design displays to achieve their desire effect (e.g., improved weather evaluation) while *also* minimizing any detrimental impact they have on decision making (i.e., induced risk taking). It should be noted, however, that the impact that advanced cockpit displays have on decision making and risk-taking behavior is likely to be affected by individual differences in pilot personalities and experiences. As stated previously, flight experience (including overall flight time, cross country flying, and recency), self-confidence, and risk-taking tendencies can all influence pilots' weather-related decision making (Goh & Wiegmann, 2002). Hence, more research is needed to examine the impact that advanced displays have on decision making in the GA cockpit, while also considering individual differences in pilots' experiences and risk-taking tendencies.

Conceptual Framework

In order to help guide further research and integrate the various issues affecting weather-related pilot decision making, a conceptual framework is presented in Figure 1. When encountering adverse weather in flight, a number of factors can influence a pilot's decision to continue flight into the deteriorating conditions. The preconditions that may affect decision making include pre-flight planning procedures and pilot characteristics such as confidence in their ability and risk taking tendencies. Inadvertent flight into IMC can be facilitated by pilots' poor situation assessment or in-flight planning or by distraction. Intentional flight into IMC could be seen as a result of pilots' low risk perception, personal motivation or perhaps social pressures. After entering into IMC, pilots typically have little time before the effects of spatial disorientation can produce catastrophic consequences.

Both technology-centered and human-centered interventions have the potential to affect pilots' weather related decision making at different stages in this model. GWIS displays for example, afford improved situation awareness and give the pilot another resource for in-flight planning. However, the use of this technology must also be considered in conjunction with pilots' self-confidence in their abilities, tendencies for risk-taking and perceptions of risk. Consequently, human-centered interventions like risk-management training need to be considered in helping reduce incidences of VFR flight into MC.

In addressing the issue of spatial disorientation, SVS displays could provide an intuitive tool for pilots to remain spatially orientated and avoid LVLOC accidents after entering IMC. However, effective training would need to be in place to help ensure the technology is used for its intended purpose (assisting in executing a 180° turn out of IMC) rather than as a means to support continued flight into conditions the pilot is not trained or qualified to fly in.

Ongoing Research

A study is currently in progress at the University of Illinois that examines how GWIS and SVS displays affect pilots' weather-related decision making. Of specific interest is how the particular properties of these two technological interventions (e.g. SVS display with and without HITS) influence pilots' decisions to continue simulated flight into deteriorating weather. In addition, the interaction of pilot personality factors with the technology will be examined. The results of the study will have implications for the design of displays for improving weather-related decision making while also minimizing risk-taking behavior.

References

Alexander, A.L., Wickens, C.D., & Hardy, T.J. (2003). Examining the effects of guidance symbology, display size, and field of view on flight performance and situation awareness. *Proceedings of the 47th Annual Meeting of the Human Factors & Ergonomics Society*. Santa Monica, CA: HFES.

Berringer, D.B. (2000). Development of highway-in-the-sky displays for flight path guidance: History, performance, results, guidelines. *Proceedings of the IEA 2000/HFES 2000 Congress*. Santa Monica, CA: HFES.

Beringer, D.B., & Ball, J.D. (2003). Effects of nexrad graphical data resolution and direct weather viewing on pilot judgments of weather severity and willingness to continue flight. *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 26-30). Santa Monica, CA: Human Factors and Ergonomics Society.

Burian, B., Orasanu, J., & Hitt, J. (2000). Weather-related decision errors: Differences across flight types. *Proceedings of the 14th IEA Triennial Congress of the International Ergonomics Association/44th Annual Meeting of the Human Factors and Ergonomics and Society* (pp. 22-24). Santa Monica, CA: HFES.

Goh, J. & Wiegmann, D. A. (2001). Visual flight rules flight (VFR) into adverse weather: An empirical investigation of the possible causes. *The International Journal of Aviation Psychology, 11*(4), 259-379.

Goh, J. & Wiegmann, D.A. (2002). Human factors analysis of accidents involving visual flight rules flight into adverse weather. *Aviation, Space, and Environmental Medicine, 73,* 817-822.

Goh, J. & Wiegmann, D.A. (2004). Effects of simulated VFR flight into IMC on pilots' confidence and perceived risk. *Proceedings of the 74th Annual Meeting of the Aerospace Medical Association*, Alaska.

Iani, C. & Wickens, C.D. (2004). Factors affecting task management in aviation. *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: HFES.

Kahneman, D. & Tversky, A. (1982). Choices, values and frames. *American Psychologist*, 39 (4), 341-350.

Klein, G.A. (1993). A recognition-primed decision (RPD) model of raid decision making. In G. A. Klein**,** J. Orasanu, R. Calderwood, & C. E. Tzambok (Eds.), *Decision making in action: Models and methods* (pp. 138-147). Norwood, NJ: Ablex Publishing Corp.

Knecht, W., Harris, H. & Shappell, S. (2004). *The influence of visibility, cloud ceiling, financial incentive and personality factors on general aviation pilots' willingness to take off into marginal weather*. Preliminary copy of FAA technical report, Civil Aerospace Medical Institute.

Latorella, K.A. & Chamberlain, J.P. (2002). Tactical vs strategic behavior: General aviation piloting in convective weather scenarios. *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: HFES

National Transportation Safety Board. (1989). *Safety report: General aviation accidents involving visual flight rules flight into instrument meteorological conditions* (NTSB/SR-89/01). Washington, DC.

Novacek, P.F., Burgess, M. A., Heck, M. L., & Stokes, A. F. (2001). *The effect of ownship information and nexrad resolution on pilot decision making in the use of a cockpit weather information display* (NASA/CR-2001-210845). Hampton, VA: NASA Langley Research Center.

O'Hare, D. (1990). Pilots' perception of risks and hazards in general aviation. *Aviation, Space and Environmental Medicine*, 61, 599-603.

O'Hare, D. & Owen, D. (1999). *Continued VFR into IMC: An empirical investigation of the possible causes: Final report on preliminary study*. Unpublished manuscript, University of Otago, Dunedin, New Zealand.

O'Hare, D., Owen, D. & Wiegmann, D. A. (2001). The "where" and the "why" of cross-country VFR crashes: Database and simulation analyses. *Proceedings of 45th Annual Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA: HFES.

O'Hare, D., & Smitheram, T. (1995). "Pressing on" into deteriorating conditions: An application of behavioral decision theory to pilot decision making. *The International Journal of Aviation Psychology, 5*, 351-370.

Prinzel, L.J., Consock, J.R., Glaab, L.J., Kramer, L.J., Arthur, J.J.& Barry, J.S. (2004). The efficacy of head down and head up synthetic vision display concepts. *International Journal of Aviation Psychology*, 14 (1), 53-77.

Takallu, M.A., Wong, D.T. & Uenking, M. D. (2002). *Synthetic vision systems in GA cockpit—evaluation of basic maneuvers performed by low time GA pilots during transition from VMC to IMC*. International Advanced Aviation Technology Conference. Anchorage, AK.

Wiegmann, D., Goh, J., & O'Hare, D. (2002). The role of situation assessment and experience in pilots' decisions to continue visual flight rules (VFR) flight into adverse weather. *Human Factors, 44*(2), 189-197.

Wickens, C.D, Goh, J., Helleberg, J., Horrey, W. & Talleur, D.A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors,* 45 (3), 360-380.

Wickens, C.D., Ververs, P.M. & Fadden, S.F. (2004). Head up displays. In D. Harris (Ed), *Human Factors for Civil Flight Deck Design*. United Kingdom: Ashgate Publishing.

Wiggins, M., & O'Hare, D. (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology: Applied, 1*, 305-320.

Williams, K.W. (2002). Impact of aviation highway-in-the-sky displays on pilot situation awareness. *Human Factors*, 44, 18-27.

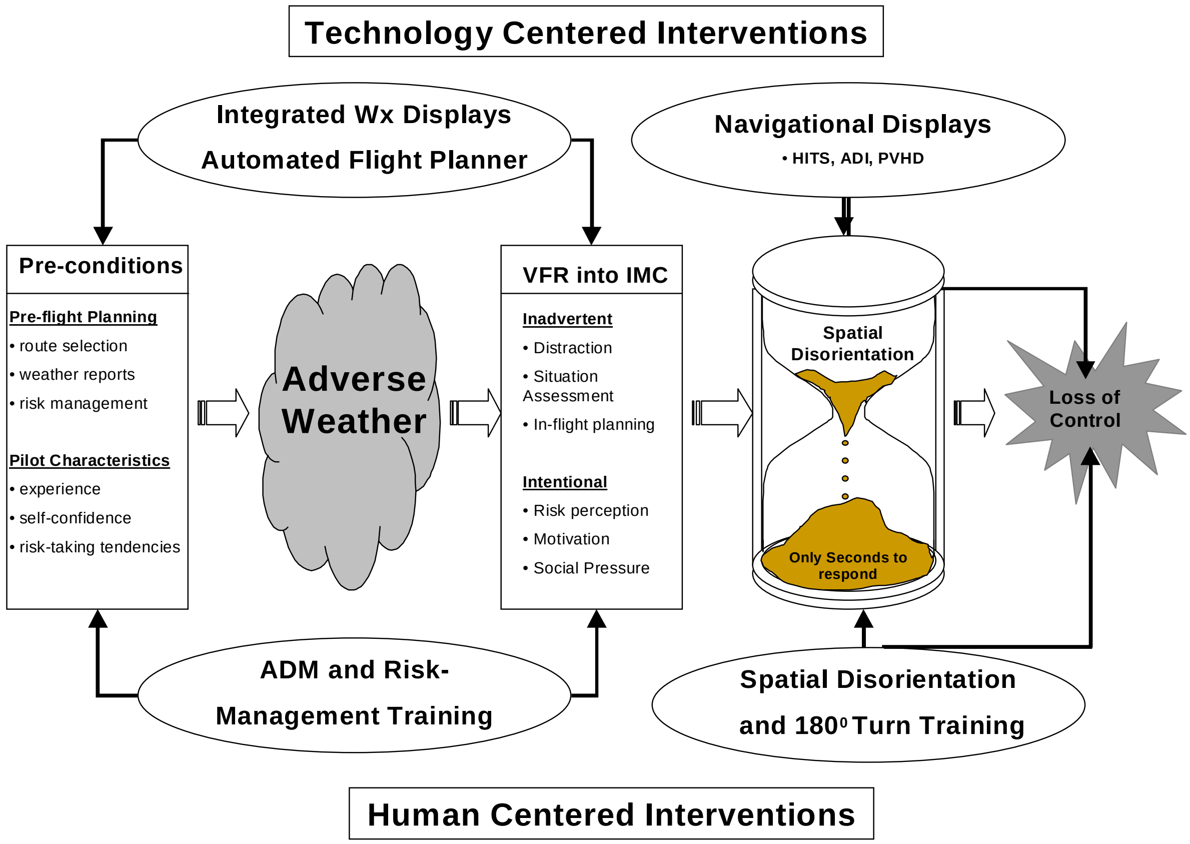


Figure 1. *Conceptual framework for integrating issues relevant to pilots' weather related decision making.*

AN ANALYSIS OF COMMUNICATIONS FOR ARRIVAL IN REAL-TIME AIR TRAFFIC CONTROL SIMULATION

Kota Kageyama
Hisae Aoyama
Electronic Navigation Research Institute
Tokyo, Japan

Air Traffic Control (ATC) communications between air traffic controllers and pilots are the most essential task in ATC operations. This paper describes results from an analysis of ATC communications in real-time (human-in-the-loop) simulation experiments. We analyzed communication data for arrival traffic of a radar control position in a Japanese terminal airspace. We assumed that particularly in a radar position, traffic volume directly influences ATC communication volume. In order to verify this assumption, we study on correlation between traffic volume and the amount of communication events issued by controllers. In addition, we examined ATC clearance phraseology and evaluated as a potential indicator of controller workload level.

Introduction

Significant airport expansion and construction programs are currently underway in Japan in order to meet the expected increase in air traffic demand. In every such airport expansion and construction project, airspace capacity estimation is one of the most important tasks. Tofukuji has asserted that air traffic controllers workload is critical to the airspace capacity, since they are directly involved in central ATC function such as decision making for control, ATC clearance issuance to pilots (Tofukuji 1993). Because it is easily observable, the amount of controller-pilot ATC communication event is often regarded as an indicator of controller workload.

Analyzing communication data from real-time ATC simulation experiments, we examined effect of traffic volume on controller-pilot communication amount. Controller-pilot communications have been assumed to increase as the function of traffic volume and other parameters (Manning *et al.* 2003). We analyzed communication data for arrival traffic of a radar control position in a Japanese terminal airspace.

In addition, we studied the effect of traffic volume and ATC communication event count on word omission in controller phraseology in order to examine the potential applicability of ATC clearance phraseology as an indicator of controller workload.

Data Acquisitions

Simulation System

The Electronic Navigation Research Institute (ENRI), a Japanese research institute for ATC systems and air navigation aids, has developed a large-scale real-time ATC simulation system for terminal and enroute radar ATC. The simulation system includes eight (8) radar displays for terminal ATC as well as several pseudo-pilot consoles.

During simulation experiments, all the communication events are recorded onto Magnet-Optical (MO) disk media by the simulation system. In order to issue communication events, controllers and pseudo-pilots press a push-to-talk button on the Plantronics® amplifier of their handsets. The time points at which the button is pressed (start-time) and released (end-time) are captured for each communication event. The recorded time points are then used to provide an efficient estimation of communication time.

The simulation system also has the ability to record aircraft trajectory data such as the temporal transition of position, altitude, speed and radar control position for each (pseudo) aircraft.

The Modeled Terminal Airspace

We conducted a series of real-time ATC simulation experiments for Japanese terminal airspace that was modeled in the simulation system and full performance level (FPL) controllers from the terminal radar ATC facility participated in the experiments. Figure 1 depicts the boundary line of the modeled airspace and the corresponding arrival flow, which is represented by arrows. The modeled airspace covers for 60NM radius from the airport up to 17,000 ft in altitude.

Depending on the flight phase, ATC arrival operations were divided into multiple radar positions. One radar position sequenced arrivals coming in from multiple directions and then, control of each arriving aircraft was handed off to other radar position that guided them through their final approach courses and assured the required separation between them.

In this paper, we focused on the radar position for final approach guidance. The oval in Figure 1 represents an example of area covered by the focused radar position.

Three (3) types of final approach course (FAC) were simulated, which, for the reminder of this paper, will be denoted as *FAC-A*, *FAC- B*, *FAC-C* respectively. We conducted four (4) experiments for each FAC type. Depending on the simulated FAC type, the experiments will be referred to as Experiment A1~A4, Experiment B1~B4, Experiment C1~C4.

An identical traffic scenario was applied to all experiments and each experiment was conducted for 75 minutes. In each experiment instance, the controllers rotated their assignment of radar positions once and therefore, in total six (6) controllers were assigned to the focused radar position throughout all experiments.

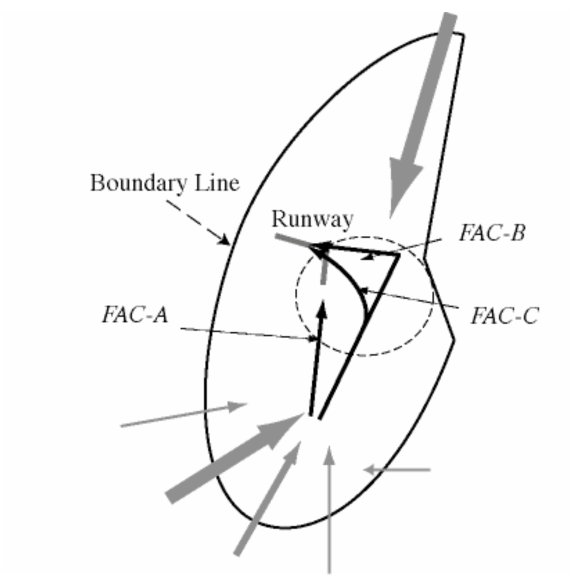


Figure 1. *Terminal Airspace Boundary Line and Corresponding Arrival Flow*

Communication Data

During the experiments, communication events between the controllers and pseudo-pilots were audio-recorded. Then, the audio data was transcribed to communication data, which contains start-time, end-time and contents for each communication event. Since these time points were measured based on the manipulations of push-to-talk buttons, there was always a possibility that the values were not as precise as the ones based on audio recordings. The start-time and end-time were recorded on the second time scale. The contents of each communication event were verbatim transcription of each audio-recorded communication event.

Data Analysis

Hypotheses

Regarding the effect of traffic volume on ATC communication events at the focused radar position, the following two hypotheses were made.

- Communication event count / time amount of communication event and traffic volume are positively correlated.
- Frequent communication events or heavy traffic volume incurs word omission.

The idea behind the first hypothesis is that heavy traffic volume triggers frequent communication events. The amount of Controller-pilot communication has been assumed to increase a) as a function of the traffic volume, and b) in airspace that utilizes complex procedures (Manning *et al.* 2003). Since the data analysis was performed on data from a radar position that was entirely dedicated to guiding arriving aircraft in their final approach courses, the airspace utilization procedures were not very complex and therefore, communication event count was assumed to be influenced only by traffic volume. Because the particular design of a FAC is also expected to influence the amount of communication events, the hypothesis was examined for each of the three types of FAC mentioned earlier.

In the second hypothesis, focus was centered on phraseology in ATC clearances. It was realized that, in some cases, controllers omit some words in their phraseology. Therefore, initially, word omission was assumed to occur frequently when controllers operate under high workload conditions and if the assumption was finally verified by the data analysis, the word omission count could then be used as an indicator of controller's response to high workload.

Analysis Methodology

After analyzing the ATC communication data issued at the focused radar position that covers the vicinity of the airport and the arrival trajectory data, the two hypotheses were verified. For the data analysis computations, each experiment was divided into fifteen (15) 5-minute time bins. These bins were combined across all four experiments for each FAC type and therefore, each FAC produced (4*15 =) 60 bins of data. For each bin, the following items were obtained and correlation between these items was calculated.

- Communication event count
- Time amount of communication events
- Word omission count
- Traffic volume

The duration of each communication event was computed by subtracting the event start-time from its end-time. Time amount of communication events in each time bin was accumulation of the duration. Because start-time and end-time were recorded on the second time scale, the time amount was also computed on the second time scale.

Word omission count was obtained by inspecting the contents of the communication events.

Traffic volume was calculated as the average number of controlled aircraft in each time bin:

$$Trafficvolume = \frac{\sum_i t_i}{\Delta T},$$

where ΔT represents length of the time bin (in this case, 5 minutes) and t_i represents the time periods during which each arrival was under the control of the focused radar position in the corresponding time bin.

Simulation Results

Communication

Table 1 shows the average and maximum values for communication event count and time amount of communication events across the 5-minute (300-second) time bins. In peak periods, the controllers issued ATC communication events approximately 25 times.

Table 1. *The Average and the Maximum Value for Communications*

FAC	Count		Time Amount (sec.)	
	Ave.	Max.	Ave.	Max.
A	12.6	23	56.8	104
B	15.7	28	66.2	114
C	16.1	25	65.4	118

Table 2 shows the correlation between the count and the time amount of communication events. As Manning *et al.* showed and as it is evident by the R^2 values in the table, they are highly correlated (Manning *et al.* 2001).

Table 2. R^2 *Values between Count and Time Amount*

FAC	Value
A	.95
B	.91
C	.96

The total duration of communications between the controllers and the pseudo-pilots was also examined and the results are presented in Table 3, where the duration of communication events issued by both the controllers and the pseudo-pilots are combined. The average percentage of the bin time spent for the communications ranged from 38.8%(*FAC-A*) to 44.1%(*FAC-B*). In the peak periods, the same percentage ranged from 61.3%(*FAC-A*) to 70.1%(*FAC-B*).

Table 3. *Time Amount for Communications between the Controllers and the Pseudo-pilots (sec.)*

FAC	Ave.	Max.
A	116.4	184
B	132.5	212
C	120.1	195

Word Omission

Table 4. *Percentage of Word Omission*

FAC	Percentage
A	4.1
B	5.5
C	5.9

Table 4 shows the percentage of word omission (WO) for the total communication count. For example, in *FAC-A*, 31 word omissions were detected whereas the total communication count was 758. The percentage is thus calculated as (31/758=) 4.1 %.

Some examples of the detected omission follow (the phrases in the parentheses represent prescribed phraseology for each case):

- “Heading 290” (Turn right/left heading 290)
- “Speed 230” (“Reduce speed to 230 knots”)
- “Contact Tower 1181” (“Contact Tokyo Tower 118.1”)
- “Approach” (“Tokyo Approach”)

Some controllers issued all their corresponding ATC clearances in using the same form of WO, in which case the omissions were considered habitual and therefore were excluded from the omission detection analysis. It should be noted that the omission percentages (as shown in Table 4) were rather low and even though some words were omitted, the instructions were sufficiently understandable for pseudo-pilots.

Traffic Volume

Descriptive statistics for traffic volume across time bins are shown in Table 5. Regardless of the value of the Standard Deviation (SD), the maximum values were more or less at the same level, because the controllers that handed off control of arriving aircraft to the focused radar position regulated the arrival transfer volume.

Table 5. *Descriptive Statistics for Traffic Volume*

FAC	Ave.	SD	Max.
A	5.9	1.9	9.4
B	6.9	2.4	9.2
C	6.6	2.7	9.4

Examination of Hypotheses

The First Hypothesis

To examine the first hypothesis, the correlation between traffic volume and communication event count / time amount of the communication events were computed for each FAC type. Table 6 shows these correlation results. The count and the time amount are both correlated with traffic volume for all FAC types, which confirms that they were both influenced by traffic volume. Although the level of the correlation varied slightly from on FAC type to another, there seems to be a stronger correlation level in the case of *FAC-C*.

To investigate the difference in correlation amongst the FAC types, the contents of the communication events were examined. The communication events were parsed into communication elements and these elements were classified according to their purpose (Prinzo 1997). Then, the frequency of principal ATC clearance elements for altitude, speed and heading were compared.

Table 6. *Correlations for Traffic Volume*

FAC	Count	Time Amount
A	.66	.62
B	.76	.74
C	.81	.77

Figure 2 represents the comparison results, which show that, compared to other types, altitude clearances were more frequently issued for *FAC-C*. This happens because, while in *FAC-A* and *FAC-B* the altitude clearances were issued entirely for the predetermined altitude of the final approach courses, in *FAC-C*, in addition to clearances for the predetermined altitude, altitude clearances were also issued to ensure separation amongst the arrival aircraft. This resulted to a larger number of altitude clearances elements in the case of *FAC-C*.

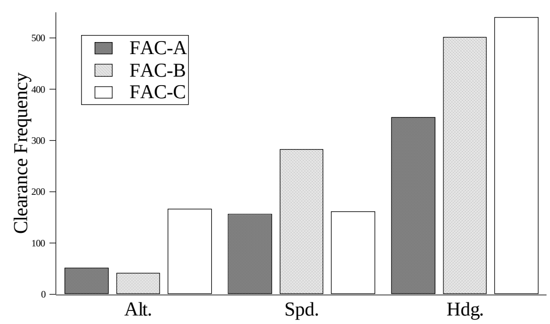


Figure 2. *Comparison of ATC Clearances*

Table 7 presents the correlation statistics between traffic volume and the principle types of ATC clearance elements. In *FAC-C*, heading and altitude clearances were strongly correlated with traffic volume.

Table 7. Correlation for ATC Cleanses

FAC	Altitude	Speed	Heading
A	.37	.49	.34
B	.27	.28	.73
C	.56	-.02	.74

The observed strong correlation of heading and altitude clearance elements to traffic volume and the higher number of altitude clearance elements observed for *FAC-C* confirmed the strong correlation between traffic volume and communication count / time amount of communication events recorded in Table 6. Furthermore, the varying frequency of the principle ATC clearances elements and the varying level of their correlation to traffic volume confirmed differences in arrival guidance methods among the three different FAC types. To further investigate the difference of influence of traffic volume among the FAC types, the guidance methodology for each FAC type should be examined and compared.

The Second Hypothesis

To examine the second hypothesis, the correlation of WO count to communication event count and traffic volume were computed and Table 8 shows the results. The superscripts "*" and "**" denote that the correlation is significant at $p < .05$ level and $p < .01$ level respectively. Except for the correlation to communication count in *FAC-B* and to traffic volume in *FAC-C*, correlations are found not to be significant. Also, the correlations are not found to be strong in any of the cases examined.

Table 8. *Correlation of WO Count*

FAC	Com. Count	Traffic Volume
A	.02	.16
B	.41**	.22
C	.22	.29*

Figure 3 represents the total count of WO for each experiment, which evidently varies a lot amongst all the experiments.

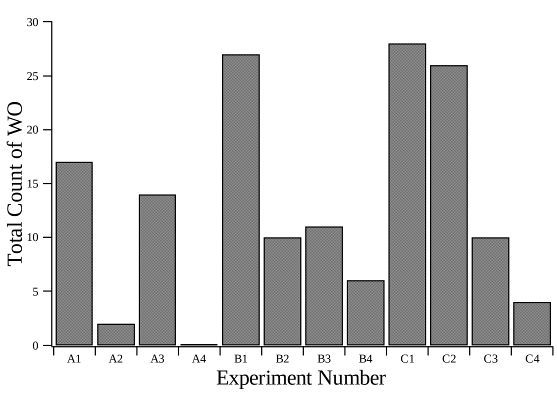


Figure 3. *WO Count on Each Experiment*

The temporal transition of WO occurrence in each experiment was also examined. The analysis results for temporal transition of WO count, communication count, and traffic volume on each experiment are shown respectively for each FAC tested in Figure 4, Figure 5 and Figure 6. Based on the FAC types, the temporal transitions curves are aligned for each experiment. The "A" to "F" below experiment numbers represent the individual controller assigned to the focused radar position during the corresponding simulation time.

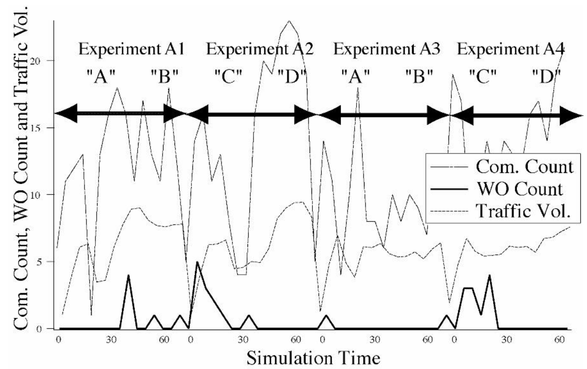


Figure 4. *Temporal Transition of WO Count for FAC-A*

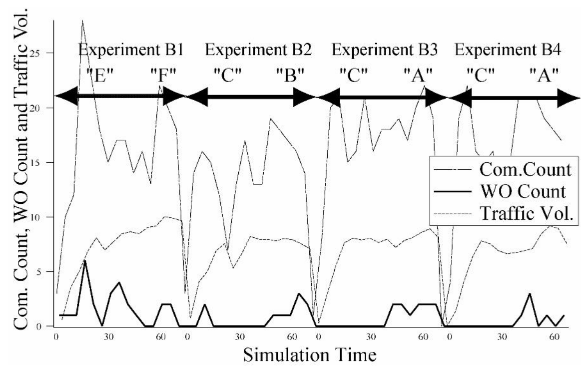


Figure 5. *Temporal Transition of WO Count for FAC-B*

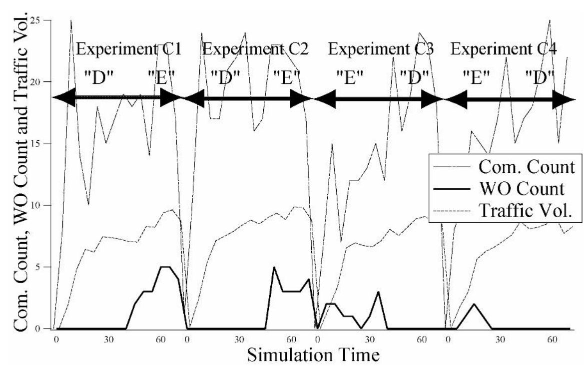


Figure 6. *Temporal Transition of WO Count for FAC-C*

It can be seen that among the different experiments, distribution of WO count is not constant and instead WO occurred frequently in certain time periods in most experiment cases. For example, while WO occurred frequently at the beginning of A2 and A4, it occurred frequently at the end of A1. In addition, WO occurrence appears to be controller-dependent in some cases. For instance, the "D" did not exhibit WO occurrence in any experiments, whereas the "E" exhibits WO occurrence repeatedly. The variation of WO count and distribution can be attributed to difference in habitual practice of individual controllers. Nevertheless, regardless of individual controller practices, a tendency for WO occurrence is observed when the controllers issued more communication events or when traffic volume was at a peak.

Summarizing, even though the hypothesis that communication event count and traffic volume influenced the WO count was not verified, there was still a possibility that frequent communication events and heavy traffic volume incurred WO in some cases.

Summary

Analyzing the communication data from real-time terminal approach radar control simulation experiments, the effect of traffic volume on ATC communication volume was studied. Analysis focus was shed on one radar control position that guides arrivals through their final approach courses and experiments were performed for three types of final approach course.

In the analysis, the correlation between traffic volume and the amount of communication events issued by controllers was examined. Although the correlation was generally strong, it appeared to be slightly different amongst the various Final Approach Course (FAC) types considered. To investigate the correlation difference among final approach course types, arrival guidance methodology must be examined and compared.

The influence of traffic volume and communication event count on word omission frequency was also examined, but no strong correlation was discovered and the omission occurrence frequency tended to be controller-dependent. On the other hand, in some cases there was a possible correlation observed between frequent communication events / heavy traffic volume and WO occurrence.

Acknowledgements

Many thanks must be attributed to Dr. Ioannis Anagnostakis for his expert advices. The authors would like to thank the controllers that participated in the simulation experiments. The authors also wish to thank the members of the ATC section at the Electronics Navigation Research Institute (ENRI) for their hard work in conducting the simulation experiments.

References

Manning, C., Mills, S., Fox, C., Pfleiderer, E. & Mogilka, H. (2001). The Relationship Between Air Traffic Control Communication Events and Measures of Controller Taskload and Workload. In proceedings of 4th USA/Europe Air Traffic Management R&D Seminar, Santa Fe, New Mexico.

Manning, C., Mills, S., Fox, C. & Pfleiderer, E. (2003). Relationships between Measures of Air Traffic Controller Voice Communications, Taskload, and Traffic Complexity. In proceedings of 5th USA/Europe Air Traffic Management R&D Seminar. Budapest, Hungary.

Prinzo, O. V. & Pickett, E. (1998). *An Acoustic Analysis of ATC Communication.* (DOT/FAA/AM-98/20). Office of Aviation Medicine, Washington D. C.20591.

Prinzo, O. V. (1998). *An Analysis of Voice Communication in a Simulated Approach Control Environment.* (DOT/FAA/AM-97/17). Office of Aviation Medicine, Washington D. C.20591.

Tofukuji, N. (1993). An Enroute ATC Simulation Experiment for Sector Capacity Estimation. *IEEE Trans. on Control Systems Technology, Vol. 1, 138-143.*

EVALUATION OF AIRCRAFT MAINTENANCE OPERATIONS USING PROCESS MEASURES

Kunal Kapoor, **Pallavi Dharwada**, **Nikhil Iyengar**, **Joel S. Greenstein**, **Anand K. Gramopadhye**
Human Computer Systems Laboratory, Clemson University, South Carolina

This research focuses on the development of a proactive system (a Web-based Surveillance and Auditing Tool - WebSAT), which promotes standardization in data collection and identifies the contributing factors that impact aircraft safety. This system will document the processes and the outcomes of maintenance activities, make the results more accessible, and reduce future maintenance error rates. WebSAT will capture and analyze data for the different operations involved in surveillance, auditing, and airworthiness directives. To achieve standardization in data collection, data needs to be collected on certain variables which measure maintenance processes. These variables are defined as process measures. The process measures incorporate the response and observation-based data collected during surveillance, audits, and the control of the airworthiness directives. This paper elaborates on the processes that exist in the aviation maintenance work group, the concerns that need to be addressed while identifying the process measures, and the utility of these process measures in conducting data analysis. Once data is captured in terms of these process measures, data analysis can be conducted to identify the potential problematic areas affecting the safety of an aircraft.

Introduction

The mission of the FAA is to provide safe and reliable air transportation and to ensure aircraft airworthiness. Maintenance error has been found to be a crucial factor in aircraft accidents (Boeing/ATA, 1995). The increasing number of maintenance and inspection errors in the aviation industry has motivated the need for human factors research. Human factors research in maintenance has deemed the human as the central part of the aviation system (Gramopadhye et al., 2000). The emphasis on the human and his role in aviation systems results in the development of error tolerant systems. Such systems will be efficient if they closely monitor and evaluate aircraft maintenance and inspection activities. Air transportation is becoming increasingly complex. The significance of the maintenance function was captured by Weick et al. (1999) when they observed that: “Maintenance people come into contact with the largest number of failures, at earlier stages of development, and have an ongoing sense of the vulnerabilities in the technology, sloppiness in the operations, gaps in the procedures, and sequences by which one error triggers another”. Given the ever increasing complexity of aircraft, a significant proportion of these errors come at the hands of the maintenance personnel themselves, due to greater demands on these individuals. Thus, it is very important to take a closer look at the humans involved in aviation maintenance, understand the causal factors for their errors and the possible solutions to counter this situation.

The aviation maintenance industry has also invested a significant effort in developing methodologies for investigating maintenance errors. The literature on human error has its foundations in early studies of errors made by pilots (Fitts, 1947), work following the Three Mile Island incident, recent work in human reliability and the development of error taxonomies (Swain and Guttman, 1983, Norman, 1981, Rouse and Rouse, 1983, Rasmussen, 1982, Reason, 1990). This research has centered on analyzing maintenance accidents. Figures emerging from the United Kingdom Civil Aviation Authority (CAA) show a steady rise in the number of maintenance error mandatory occurrence reports over the period 1990 to 2000 (Courteney, 2001). A recent Boeing study of worldwide commercial jet aircraft accidents over that same period shows a significant increase in the rate of accidents where maintenance and inspection were primary factors (ICAO, 2003). The FAA, in its strategic plan for human factors in aviation maintenance, through to 2003, cited statistics from the Air Transport Association of America (ATA) showing that the number of passenger miles flown by the largest US airlines increased 187% from 1983 through to 1995. Over that same period, the number of aircraft operated by those airlines increased 70%, but the number of aviation maintenance technicians increased only 27%. The FAA concluded that the only way the maintenance program could cope with the increased workload was by increased efficiency at the worker level (McKenna, 2002).

Attempts have been made to define a core set of constructs for a safety climate (Flin et al., 2000). Although not entirely successful in establishing core dimensions, this research is useful in suggesting constructs that should be considered for inclusion in research on maintenance errors. Taylor and Thomas

(2003) used a self-report questionnaire called the Maintenance Resource Management/Technical Operations Questionnaire (MRM/TOQ) to measure what they regarded as two fundamental parameters in aviation maintenance: professionalism and trust. The dimension of professionalism is defined in their questionnaire in terms of reactions to work stressors and personal assertiveness. Trust is defined in terms of relations with co-workers and supervisors. Patankar (2003) constructed a questionnaire called the Organizational Safety Culture Questionnaire which included questions from the MRM/TOQ along with items from questionnaires developed outside the maintenance environment. Following the application of exploratory factor analytic routines to a dataset generated from respondents that included 124 maintenance engineers, Patankar identified four factors as having particular relevance to the safety goals of aviation organizations. They are emphasis on compliance with standard operating procedures, collective commitment to safety, individual sense of responsibility toward safety, and a high level of employee-management trust.

In addition to descriptive accident causation models, classification schemes, and culture surveys, there is a need for empirically validated models/tools that capture data on maintenance work and provide a means of assessing this data. However, such models and schemes often tend to be ad hoc, varying across the industry, with little standardization. In order to contend with this issue, new empirical models and tools are needed which employ standardized data collection procedures, provide a basis for predicting unsafe conditions, and design interventions that will lead to reductions in maintenance errors.

Process Measures

This research seeks to identify error causes and occurrences using a web based surveillance and auditing tool (WebSAT). The purpose of WebSAT is to capture and analyze data for different processes involved in the surveillance, auditing, and airworthiness directives functions of the aviation maintenance industry. To achieve standardization in data collection, data needs to be collected on certain variables which measure maintenance processes. These variables are defined as process measures.

The process measures incorporate the response and observation-based data collected during surveillance, audits, and the airworthiness directives control processes. Once data is captured in terms of these process measures, data analysis can be conducted to identify the potential problematic areas affecting the safety of an aircraft. In this stage of data analysis, the performance of processes and those conducting these processes will also be evaluated.

Quality Assurance Work Functions

The complexity of the inspection and maintenance system is complicated by a variety of geographically dispersed entities ranging from large international carriers, repair and maintenance facilities through regional and commuter airlines, to the fixed-based operators associated with general aviation (Kapoor et al., 2004, Dharwada et al., 2004). Inspection is regulated by the FAA, as is maintenance. However, while adherence to inspection procedures and protocols is closely monitored, evaluating the efficacy of these procedures is much more difficult. This section explains the quality assurance work functions which are responsible for aircraft maintenance.

Surveillance

Surveillance is the day-to-day oversight and evaluation of the work contracted to an airframe substantial maintenance vendor to determine the level of compliance with airline's Maintenance Program and Maintenance Manual with respect to the airline's and FAA requirements. For example, FedEx, our partner in this project has a surveillance representative, stationed at the vendor location who schedules surveillance of an incoming aircraft. The specific task to be performed on an aircraft at a vendor location is available on a work card. The representative performs surveillance on different work cards according to a surveillance schedule. The results are documented and used to analyze the risk factors associated with the concerned vendor and aircraft. The FedEx surveillance department classifies the data obtained from a surveillance visit at the maintenance facility into categories. These categories are based on various surveillance tasks and the C.A.S.E. (Coordinating Agency for Supplier Evaluation) guidelines that are adhered to by the substantial maintenance vendor and the airline. The team used these categories as a starting point to identify process measures. Some of the categories currently being used by FedEx are in-process surveillance, final walk around, and verification surveillance.

Technical Audit

The system level evaluation of standards and procedures of suppliers, fuel vendors, and ramp operations done on a periodic basis is referred to as Technical Audit. The work function of technical audits is to ensure compliance with Federal Aviation Regulations (FARs), and established company policies and procedures. The team worked towards identifying process measures for this work function. Data collected from the technical audit checklists will be utilized for analysis on the effectiveness of the technical audit process.

Internal Audit

The evaluation of internal processes in the departments of an airline is referred to as Internal Audit. The work function of the internal audit department is to sample the processes being used by departments in an organization and to verify their compliance with regulatory, company and departmental policies and procedures. Similar to the technical audits, the data collected from internal audit checklists will be grouped into process measures to facilitate further data analysis and assess the effectiveness of the internal audit process.

Airworthiness Directives Department

The evaluation of the applicability, loading, and tracking of airworthiness directives is referred to as airworthiness directives control. The work function of the Airworthiness Directives (AD) control department is to review AD-related Engineering Order/Work Instruction Cards (EO/WIC), the acquisition process, and the customer's maintenance manual. The data collected from these processes will be grouped into categories to facilitate further data analysis and assess the effectiveness of the airworthiness directives control department.

Observations during the Identification of the Process Measures

The team adopted the following data collection methods: Interviews, Observation Sessions, Document Study, and Questionnaires (Iyengar et al., 2004). The team determined that the process measures being identified must include all the data that is gathered during the maintenance operations. The team observed inconsistency in the definition of the existing categories among the surveillance representatives. The representative's own experience could be a road block, preventing him from correctly assigning an error to a category. The internal audit department employed a definitive structure of six categories, and after scrutiny of the internal audit documents, the team concluded that these categories covered the entire span of the data generated during audits in the internal audit department. The data analysis in the technical audit department lacked strategy. The personnel in the airworthiness directives department utilized canned statements for data analysis, which lacked strategy. There were two major work domains being considered in the AD department: information verification based on AD department-related engineering order/ work instruction cards (EO/WIC), manuals and other documents involved with the compliance of airworthiness directives. The AD department also verifies information related to AD status reports.

Observations for Surveillance

The surveillance representatives relied on their memory to categorize what they saw in the maintenance facility. This suggested that there must be a manageable number of categories and they should be easy to remember. There were process measures being used for data analysis in surveillance, some of which were redundant, and there was no consensus among the surveillance personnel within the department at FedEx in the classification of a work card into a specific process measure. There were two distinct categories of process measures: Technical and Non-Technical. Process measures which include surveillance involving scheduled maintenance activities performed on an aircraft during a maintenance event are referred to as technical process measures. These process measures include technical activities that are hands-on and performed directly on the aircraft. Technical activity also includes maintenance that is performed in a back shop setting on a removed aircraft part. An example would be a panel removed and routed to a composite back shop for repair, then reinstalled on the aircraft. The surveillance activities involving verification of standardized procedures, referenced manuals, equipment, and facility maintenance requirements are referred to as non-technical process measures. It was important for the team to understand the purpose of the data being gathered and its relevance to aircraft safety. Hence, collection of data on non-technical measures was given equal emphasis on technical measures. The team recognized the importance of incorporating the concerns of the quality assurance representatives while finalizing the list of process measures for surveillance.

Observations for Internal Audits

The internal audit department at FedEx was working with a robust set of process measures. These were administration, training, records, safety, manuals, and procedures. The team scrutinized the documents and check lists the personnel in the internal audit department work with. These process measures would effectively categorize all the data being generated in this department.

Observations for Technical Audits

The technical audits department conducts annual audits on all FedEx vendors. These vendors are substantial supplier vendors, fuel, ramp operations, and aircraft maintenance vendors using checklists which are query based. The team determined that each check list had a series of questions dedicated to one fundamental domain, such as inspection or facility control. These domains were consistent for the different checklists emphasizing the needs of diverse vendors such as the supplier vendor and the fuel vendor. A final consensus within the research team finalized the process measures as these categories within check lists itself.

Observations for Airworthiness Directives Department

The personnel in this department are involved in two primary activities. They validate the information presented on AD-related EO/WIC, manuals, status reports and other documents involved with the compliance of airworthiness directives. The personnel also verify the adequacy of the activities involved in the loading and tracking of airworthiness directives, including inspection intervals.

Process Measures Validation

Once the research team finalized the process measures definition document, and finalized a list of the process measures to be used for the different work functions, it was important for the research team to validate their research efforts. The team conducted a two-phase on-line survey to validate results. The online survey was initially sent to the surveillance, auditing, and airworthiness directives department personnel at FedEx. There were six participants from each department. Prior to the participants taking the survey, the research team sent out an e-mail to them. This e-mail had detailed instructions about how to take the survey, and the team also expressed the goal of the survey. A process measure definitions document to be read before taking the survey was sent to the participants. The survey had four modules. The survey was designed to last a maximum of 60 minutes. It included 7 to 21 questions depending on the survey module. The questions were of two kinds. There were forced-choice questions, and open-ended questions. Each question had a field for the comments of the personnel taking the survey. The reason for this was that the team wanted detailed feedback from the participants. The participants taking the survey were not identified. The team gave two weeks to get inputs from the participants of the survey. Once the data was generated and analyzed, the research team iterated its definition document to incorporate changes expressed by the participants.

In the next phase, the research team sent out the same survey to other supporting and partnering airline organizations: Alaska Airlines, Delta Airlines, IATA, and America West. The results of this survey are still awaited.

Use of Process Measures in WebSAT

The following is a list of identified process measures for the fours modules WebSAT is involved with.

Process Measures for Surveillance
1. In process Surveillance
2. Verification Surveillance
3. Final Walk Around
4. Documentation Surveillance
5. Facility Surveillance
6. Procedures Manual Surveillance

The other data capturing modules in surveillance which facilitate capturing of the data but are not process measures of the surveillance work function are given below:

1. Additional Findings Module
2. Fuel Surveillance Module

The above mentioned modules are not process measures since they do not evaluate the routine surveillance process. The information captured from the additional findings module is important for an airline for documentation purpose. This data is not used to rate vendor performance of maintenance tasks. Fuel surveillance is not performed in every maintenance facility. To avoid inconsistencies in data classification across the facilities, the team proposed to treat the process of fuel surveillance as a separate module. The data captured in this module will be analyzed separately to comment on the effectiveness of fuel surveillance.

Process Measures for Internal Audits
1. Administration
2. Training
3. Records
4. Safety
5. Manuals
6. Procedures

Process Measures for Technical Audits
1. Compliance/ Documentation
2. Inspection
3. Facility Control
4. Training and Personnel
5. Procedures
6. Data Control
7. Safety

Process Measures for Airworthiness Directives
1. Information Verification
2. Loading and Tracking Verification

The WebSAT framework strategy for the research revolved around three tiers (stages). The first tier involved the collection of data with respect to work functions of surveillance, auditing (internal & technical), and airworthiness directives. Once the data involving the maintenance of an aircraft was gathered from these sources, they would be scrutinized with respect to the process measures. In the next stage, tier 2, the analysis of the relevant data would be categorized. In tier 3, a final analysis would categorize the variables into risk (impact variables), and non-risk variables. To implement this framework, WebSAT will use a data model to interpret and analyze the data gathered. Traditional analytical techniques deal mainly with the identification of accident sequence and seek unsafe acts or conditions leading to the accident. Such techniques include the sequence of events (domino effect), known precedents etc. For example, Pate-Cornell (1993) has developed an analytical framework, to establish the causal relationship between the basic events, decision and actions, and organization factors. She demonstrated the use of this framework in the analysis of the Piper Alpha accident which occurred due to a massive explosion on the offshore oil and gas production platform (Pate-Cornell, 1993, Cojazzi and Cacciabue, 1994). However, the post-hoc nature of these frameworks renders them inadequate for a proactive WebSAT. The team hopes to develop a data model in which the process measures can be used to establish causal relationships in the QA processes.

Acknowledgements

This research is supported by a contract to Anand K. Gramopadhye and Joel S. Greenstein, Department of Industrial Engineering, Clemson University from the Federal Aviation Administration (Program Manager: Dr. William Krebs, AAR-100). Our special thanks to Jean Watson and William Krebs from FAA for extending their support in conducting this research. We would also like to thank Rocky Ruggieri, Ken Hutcherson and the Quality Assurance department team from FedEx for their cooperation in providing data and their contribution in data gathering and interpretation sessions. The opinions, findings, conclusions and recommendations presented in this paper are those of the authors and do not necessarily reflect the views of the FAA.

References

Boeing/ ATA (1995). Industry Maintenance Event Review Team. The Boeing Company, Seattle, WA. FAA (1991). *Human Factors in Aviation Maintenance Phase1: Progress Report*, DOT/FAA/AM-91/16.

Cojazzi, G. & Cacciabue, P.C. (1994). The DYLAM Approach for the Reliability Analysis of Dynamic System. *Reliability and Safety Assessment of Dynamic Process Systems*. Edited by Aldemir, T., Siu, N.O., Mosleh, A., Cacciabue, P.C., Goktepe, B.G. Springer-Verlag Berlin Heidelberg 1994.

Courteney, H. (2001). Safety is no accident. *Royal Aeronautical Society Conference*, London, United Kingdom.

Dharwada, P., Iyengar, N., Kapoor, K., Greenstein, J.S. & Gramopadhye, A.K. (2004). Web-based surveillance and auditing tool (WebSAT): A proactive system to capture maintenance errors. *Proceedings of Safety Across High-Consequence Industries*, St. Louis.

Fitts, P.M. & Jones, R.E. (1947). Analysis of factors contributing to 460 “pilot-error” experiences in operating aircraft controls. *Memorandum Report TSEAA*-694-12. Dayton, OH: Aero Medical Laboratory, Air Material Command.

Flin, R., Mearns, K., O'Connor, P. & Bryden, R. (2000). Measuring safety climate: Identifying the common features. *Safety Science*, 34, 177-192.

Gramopadhye, A.K. & Drury, C.G. (2000). Human Factors in Aviation Maintenance: How we got to where we are, *International Journal of Industrial Ergonomics*, *26*, 125-131.

ICAO 2003. *Human factor guidelines for aircraft maintenance manual*.

Iyengar, N., Kapoor, K., Greenstein, J.S. & Gramopadhye, A.K. (2004). Data Gathering Methodologies to Identify Impact Variables in Aviation Maintenance, *Industrial Ergonomics and Research Conference*, Houston.

Kapoor, K., Dharwada, P., Iyengar, N., Greenstein, J.S. & Gramopadhye, A.K. (2004). Standardized Auditing and Surveillance of Aircraft Maintenance Operations. *Industrial Ergonomics and Research Conference*, Houston.

McKenna, J.T. (2002). Maintenance resource management programs provide tools for reducing human error. *Flight Safety Foundation Flight Safety Digest*, 1-15.

Norman, D.A. (1981). Categorization of action slips. *Psychology Review 88*, 1-15.

Patankar, M.S. (2003). A study of safety culture at an aviation organization. *International Journal of Applied Aviation Studies*, 3(2), 243-258.

Pate-Cornell, M.E. (1993). Risk Analysis and Risk Management for Offshore Platforms: Lessons from the Piper Alpha Accident. *Journal of Offshore Mechanics and Arctic Engineering*, Vol. 115, Aug 1993, pg 179-190.

Pate-Cornell, M.E. (1993). Learning from the Piper Alpha Accident: A Postmortem Analysis of Technical and Organization Factors. *Risk Analysis*, 13(2), 215-232.

Rasmussen, J. (1982). Human Errors: A taxonomy for describing human malfunction in industrial installations. *Journal of Occupational Accidents*, 4, 311-333.

Reason, J.T. (1990). Human Error. Cambridge: Cambridge University Press.

Rouse, W.B., and Rouse, S.H. (1983). Analysis and Classification of Human Error. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13, No. 4, 539-549.

Swain, A.D., & Guttman, H.E. (1983). Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications: *Final Report. NUREG/CR-1278*, SAND80-0200. Prepared by Sandia National Laboratories for the U.S. Nuclear Regulatory Commission.

Taylor, J.C. & Thomas, R.L. (2003). Toward measuring safety culture in aviation maintenance: The structure of trust and professionalism. *The International Journal of Aviation Psychology*, 13(4), 321-343.

Weick, K.E., Sutcliffe, K.M. & Obstfeld, D. (1999). Organizing for high reliability: Processes of collective mindfulness. *Research in Organizational Behavior*, 21, 81-123.

INTERFACE DESIGN TO ACTUALIZE HUMAN-MACHINE COORDINATION IN THE HIGHLY AUTOMATED COCKPIT

Daisuke Karikawa, **Makoto Takahashi**, **Masaharu Kitamura**
Graduate School of Engineering, Tohoku University
Sendai, JAPAN

Capt. Akira Ishibashi
Japan Institute of Human Factors
Tokyo, JAPAN

Although automated systems in an advanced cockpit have contributed to the enhancement of the safety, they have also increased the system's complexity which can be a cause of inappropriate situation awareness (SA). In this paper, a supportive method for enhancing SA in the highly automated cockpit has been proposed, focusing on the following two points. One is to support in grasping the situation from a broader perspective which can contribute to the detection of SA errors and to better understanding of the system's activities. The other is mitigating additional cognitive burden by supportive information. For the achievement of both an informative cockpit and the minimum additional cognitive burden, we explore the interface design for supporting SA in terms of improving information management by assisting the detection of unexpected conditions in the early stages of risky situation.

Introduction

In recent years, various automated systems have been introduced to the aircraft's cockpit, and most normal operations have increasingly being accomplished through them. Although the advanced automation has no doubt contributed to the enhancement of the aviation safety, it can also be one of the reasons for the system's increasing complexity which leads to the difficulty of maintaining appropriate situation awareness (SA) [1]. In some of the aviation accidents involving advanced aircrafts, inadequate SA was an important cause of breakdown in the human-machine coordination [2, 3].

In this study, SA in the highly automated cockpit is assumed to be divided into the following two aspects. One is grasping the state of an automated system itself. A source of difficulty in acquiring this SA is the system's internal complexity. For example, the autopilot system has over 20 modes with complex mode combinations and automatic mode transitions. Some aviation accidents indicated that the complete understanding of an autopilot system can sometimes be difficult even for the highly trained pilot [2, 4]. The other aspect is SA from a broader perspective including situation and environmental condition. In this paper, global SA is used to mean this SA with bird's eye view. Global SA is important for greater understanding of the system's activities, the projection of future state and the result, that is "Why is the system doing that?", "What will the system do next?", "What will the result of the system's activities be?"[5] This awareness is essential for achieving the effective human-machine coordination. For example, misunderstanding of the consequences of the system's activities can cause inappropriate risk perception, which may result in the delay in taking remedial actions by pilots.

For supporting the SA concerning the autopilot system itself, our research group has proposed a method to support the pilot in detecting possible deviations in mental models by providing additional information to enhance SA for the actual state of the autopilot system. A prototype information display for supporting the SA has been developed in order to demonstrate the validity of the proposed method [6].

In the present study, an improved prototype interface for supporting more global SA based on the previous approach has been proposed. In other researches, it has also been pointed out the importance of a more global SA and various supportive interfaces have been proposed [7, 8]. Although they have the potential to the enhance pilot's SA, the negative aspect of providing additional support for practical and effective use, which may lead to increased complexity of displays or pilots' cognitive overload, should also be considered.

Therefore, for satisfying both of the achievement of an informative cockpit and the mitigation of additional cognitive burden, we explore the supportive interface design focusing on the assistance of information management in the early stages of risky situations.

Basic Models

In this chapter, we discuss models and definitions which are fundamental to the present study.

Situation Awareness Error

Some researches have indicated a strong relationship between SA and mental models, the latter of which are internal models of systems and environment [9, 10]. Endsley has explained the role of the mental model as below. "This situation model

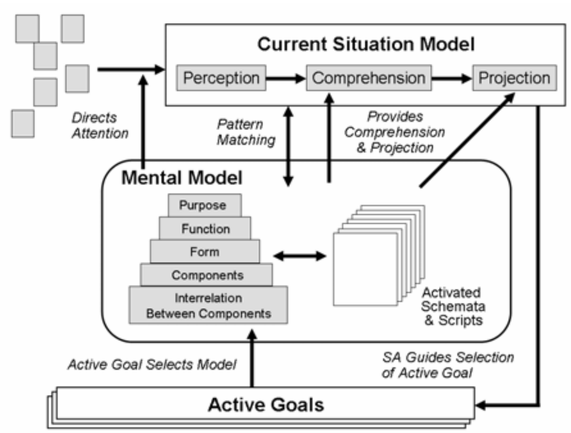


Figure 1. *Role of goals and mental models in SA [11] (M. R. Endsley, "Theoretical Underpinnings of Situation Awareness: A Critical Review", Situation Awareness Analysis and Measurement, Lawrence* Erlbaum Associates, pp.16, FIG. 1.6.)

captures not only the person's representation of the various parameters of the system, but also a representation of how they relate in terms of system form and function to create a meaningful synthesis - a gestalt comprehension of the system state" [11]. Fig. 1 shows the role of mental models in SA by Endsley. Based on the figure, the mental model affects direct attention to key features on objective system and environment for the development of SA, and this mental model is revised based on the acquired SA. It means that there is a possibility of going into a kind of error loop by the use of incorrect mental models. That is, an inadequate mental model can cause a system operator to have an erroneous SA as described in Fig. 1, and the operator might not recognize key features which might indicate the SA error because direct attention to information is controlled by the mental model updated directly or indirectly based on erroneous SA. Such a situation can be critical because it is considered to be already difficult for the operator himself to detect the SA error.

However, it also indicates the possibility that an operator can remedy the SA error himself by detecting the deviations of the activated mental model from the actual situation at an early stage of the event. In this study, our purpose is to support the detection of inaccuracy in the activated mental model by providing additional information with less cognitive burden.

Mental Model

According to Rasmussen's study, the mental model of a system operator consists of at least two dimensions which can be represented as hierarchies [12]. One is the functional hierarchy in line with the dimension of means-end. The other is the physical hierarchy along the dimension of parts-whole. In the real aircraft's operation, the mental model activated in a pilot's mind dynamically changes with the transition of active goals which are sometimes parallel or conflictive. Rasmussen stated that the physical and the functional hierarchy in a system operator's mental model dynamically interact with each other at appropriate representation levels in order to interpret and evaluate flooding information from the ongoing situation [12]. This operator's ability to describe the system at the various physical and functional levels can also provide a kind of redundancy for grasping situation. Even if an operator fails to realize the abnormal indication of gauge, it is possible that the operator can recognize the SA error by reasoning from the unexpected state of either a more global system or a lower levels system. In such a case, the operator recognizes the objective system at multi-levels of the mental models, which can provide redundant ways of recognition for the situation. Therefore, for supporting the redundant situation recognition, the information concerning the bird's eye view of a situation should be displayed at the same time as the existing indications.

Supporting Method

The importance of supporting global SA has been indicated, and new displays for supporting it have been proposed or already come into practical use, e.g. Vertical Situational Display (VSD). However, it is necessary to thoroughly consider the trade-off between their effectiveness and the cognitive burden in additional supportive information. In fact, according to a questionnaire survey of 10 commercial pilots in our previous research, many pilots were sensitive to increasing the amount of information on the interface, although they showed an interest in the possibility of additional supports to acquire the appropriate SA.

Therefore, the principle of our system design is to provide minimum additional cues on the interface for helping a pilot in detecting any inadequacy in the activated mental model in terms of the most important goal - safety. These additional cues can be provided as follows:

- Operators are NOT obligated to grasp the cues by excess highlighting or warning sound or other means. Operation procedure also does not require operators to do it.
- Operators are NOT obligated to respond to the appearance of cues.
- Cues do NOT have excess saliency compared with other displayed information.
- It should be intuitive and easy enough for operators to acquire the cues and to interpret their meaning.

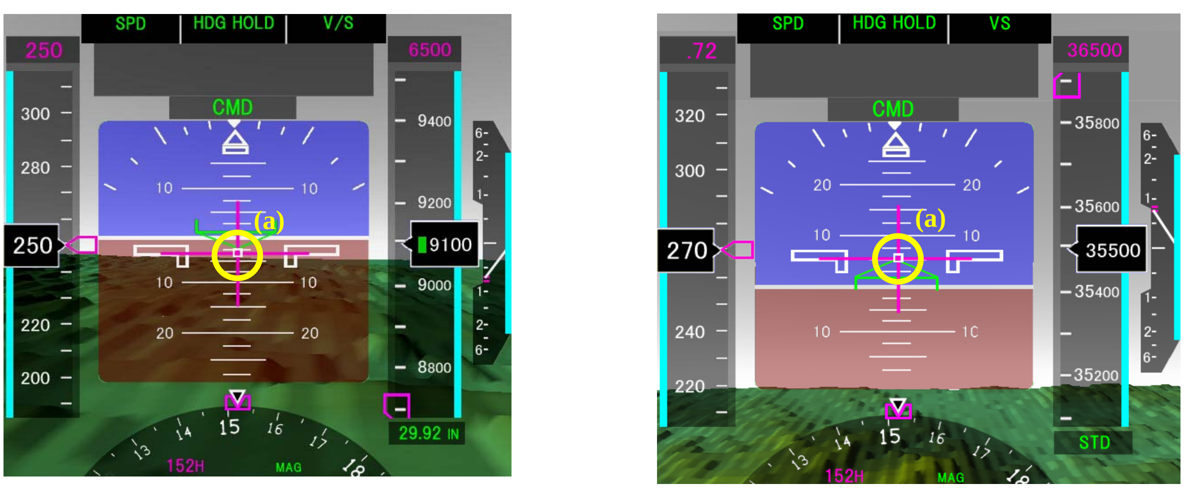
The basic concept of our method for displaying supportive information is in supporting the recognition of the deviation from expected situation in performing routine tasks in the early state of risky situation.

Implementation

Based on the discussion in previous chapters, we have proposed a prototype of the improved Primary Flight Display (PFD) for the enhancement of SA. Fig. 2 and Fig. 3 show some examples of the display. The basic structure of the display is the same as the existing PFD because pilots are highly accustomed to the existing form. Some incremental information has been added in the proposed display.

Firstly, the graphical information of the terrain is added to the proposed display. The topographic data comes from a database of existing Enhanced Ground Proximity Warning System (EGPWS). The graphical information is adjusted such that the center pointer of the aircraft symbol indicates the flight path (inside the circle (a) of Fig. 2 (1)(2)). It means that the aircraft will fly on in the direction of the center pointer. The indication can enhance the awareness for the physical relationship between the terrain, and the current position and flight path of an aircraft. As the terrain display are based on the flight path and the range of visibility is also limited to about 40km (enough to indicate terrain information for the next few minutes), the saliency of the terrain display naturally declines when the information is considered to be less important, for example, when the aircraft is climbing or cruising at the high enough altitude as described in Fig.2(2).

Compared to other terrain displays previously proposed, the saliency of the terrain indication of this display is appreciably low. Most part of it is overlaid by other indicators, giving them the priority. The reason is that, as the first aim of the terrain indication is to support the detection of possible error in the activated mental model which is used for projecting future situation, minimum information may be enough to accomplish the aim. In the context of this study, the error in the activated mental model signifies the "existence" of causes of danger which may result in a crash along the flight path which pilots do not expect, caused by some factors such as the erroneous setting of the autopilot pilot system which leads to unexpected sudden descent, or by the pilot's misunderstanding of the terrain feature. Therefore, the indication is enough to represent the relationship between the flight path and the existence of causes of a crash. There is no need for the display of detailed land features. In other words, it provides only some key features of the information of the aircraft's flight path with connection to the physical terrain information, which can describe the deviation from the pilot's expected situation more intuitively than the existing indications. This kind of global situation display can contribute to the detection of the possible deviation of



(1) descending at 9100feet

(2) climbing at 35500feet

Figure 2. *Examples of the proposed display*

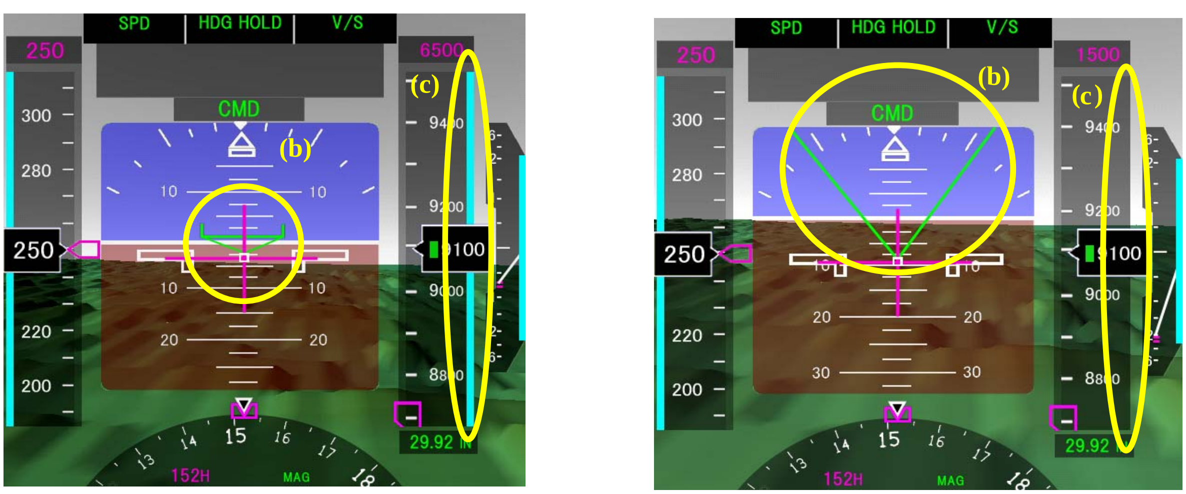
the pilot's mental model from the actual situation. This, in turn, is useful for evaluating whether the aircraft will be safe or not in the foreseeable future.

Secondly, the consequence of the activities of autopilot system is also shown in the proposed display as described inside the circle (b) in Fig. 3(1). Although the autopilot system originally consists of lower level functions or their combinations, such as maintaining the vertical speed or maintaining the heading, the proposed display represents the future situation by describing the function of the autopilot system at a more global level with connection to the physical terrain model. For a clearer indication, if the system detects a possible crash, the symbol is indicated as shown inside the circle (b) in Fig. 3(2), which is different from that in a safe situation. This indication of the intention of the autopilot system can support the pilot's more global SA. That is, the pilot can understand the situation not only at the level of "What is the system doing?", but also at the level of "What is the outcome of the system's action?" by the indication. Furthermore, if the system detects a possible crash, a blue bar which is presented on the altitude indicator turns invisible (inside the ellipse (c) of Fig. 3(1) and Fig. 3(2)). The blue bar indicates a result of safety assessment by the system, and disappearance of the indication from the display can inform the pilot about the system's abandonment of responsibility for safety when detecting a possible erroneous direction based on the system's situation assessment.

These indications described above cannot provide detailed information like the precise distance or remaining time to the possible crash. However, the proposed display can be expected to indicate possible dangers in its earlier phase by supporting the recognition of the deviation from the pilot's expected situation. In other words, the aim of the proposed display is to promote the acquisition of necessary information by the pilots in an earlier time frame, which can allow the efficient use of more supportive information like VSD.

In addition, we have explored the use of the framework of the proposed display which provides information of a more global situation for a greater understanding of the system's direction or intention especially in the critical situation. We have designed an enhanced display for TCAS based on the framework of the proposed display. Symbols of other aircraft subject to the TCAS advisory are overlaid in the proposed display described in Fig. 4. In that case, the pilot is informed of an aircraft which is coming close to the center pointer, a significant cause of danger because the center pointer indicates the flight path. The indication can support a pilot in recognizing the transition of the degree of danger intuitively. Therefore, the pilot can decide whether he/she should follow the TCAS advisory or not earlier. It can contribute to the proper understanding of the appropriateness of the TCAS advisory, and to the avoidance of erroneous decision under severe time pressure.

The proposed display can clearly indicate the possible danger of an expected crash in the early stages of the accident. The function can provide pilots with more



(1) normal descent (2) abnormal descent into the terrain

Figure 3. *Examples of the proposed display*

The aircraft is descending using the autopilot system. In Fig. 3(2), the aircraft has a possibility of crashing into the terrain because the pilot erroneously directs too low a target altitude to the autopilot system.

opportunities to recover from errors. It can help to prevent certain types of CFIT accidents such as the accident of Air Inter Flight 148 at Strasbourg in 1992, the accident of Korean Air Flight 801 at Guam in 1997, because one of the important causes of these accidents is that pilots could not recognize the situation that the aircraft was heading to the terrain erroneously. Although the evaluation of the prototype display is still ongoing, we believe that the proposed method is effective for realizing a higher level of safety in the highly automated system.

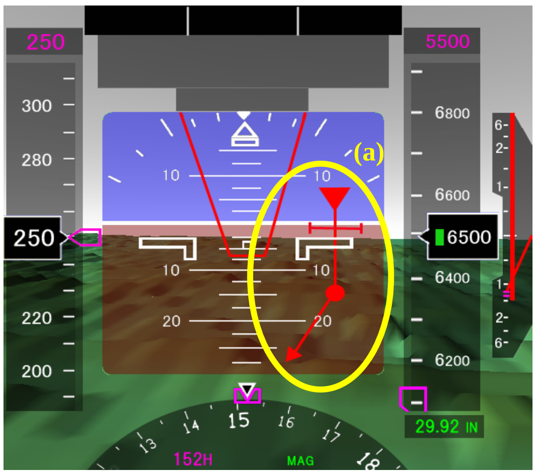


Figure 4. *Examples of the proposed TCAS display*

The symbol inside the ellipse (a) indicates the relative altitude, the position, the lateral and vertical direction of the closing aircraft.

Conclusion

We have developed the prototype interface for providing supportive information to the pilots. It aims at providing cue information for detecting possible deviation of mental model from the real situation. The great cue has been taken not to increase the cognitive burden in providing additional information in the already congested PFD. The importance point of the proposed interface point of the proposed interface concept is that it can increase the possibilities that pilot can recognize the existence of the risky situation in its early stage. Authors believe that the validity of the proposed interface concept will be validated by the experimental evaluations.

References

[1] D. D. Woods, “Tasks, Errors and Mental Models”, Taylor & Francis, London, UK, 1988.

[2] E.L.Wiener and R.E.Curry, Flight-deck automation: promises and problems, Ergonomics, Vol.23, No.10, pp.995-1011, 1980

[3] M. MacPherson, G., The Black Box, William Morrow and Company Inc., New York, U.S., 1998

[4] G.C.L.Fletcher, et.al., “Transition Training and the Glass Cockpit”, Proc. of the Ninth International Symposium on Aviation Psychology, Vol.1, pp.237-242, 1997

[5] D. D. Woods, “Decomposition Automation: Apparent Simplicity, Real Complexity”, Automation and Human Performance, Lawrence Erlbaum Associates, Mawwah, U.S., 1996, pp. 3-17.

[6] Daisuke Karikawa, Akira Ishibashi, Makoto Takahashi, Masaharu Kitamura, “Information Display for Enhancing Situation Awareness in Autopilot Systems”, Proc. of 12th International Symposium on Aviation Psychology, 2003, pp. 1306-1310

[7] Ryutaro Kawano, “Self-Object Observation Interface”, Proc. of Seminar on Advaned Human-Machine Cooperation (HMC) in Complex Dynamic System, 2000

[8] I. Solodilova, G. Lintern, N. Johnston, “The Modern Commercial Cockpit as a Multi-Dimensional, Information-Action Workspace”, Proc. of 12th International Symposium on Aviation Psychology, 2003, pp. 1096-1101

[9] N. B. Sarter & D. D. Woods, “Situation awareness: A critical but ill-defined phenomenon”, The International Journal of Aviation psychology, Vol. 1 No. 1, 1991, pp.45-57

[10] R. H. Mogford, “Mental models and situation awareness in air traffic control”, The International Journal of Aviation psychology, Vol. 7 No. 4, 1997, pp.331-342

[11] M. R. Endsley, “Theoretical Underpinnings of Situation Awareness: A Critical Review”, Situation Awareness Analysis and Measurement, Lawrence Erlbaum Associates, Mawwah, U.S., 2000, pp. 3-32.

[12] J. Rasmussen, Information Processing and Human-Machine Information? An Approach to Cognitive Engineering, Elsevier Science Publishing Company, Inc., New York, U.S., 1986

COLLECTIVE TRAINING RESEARCH UTILIZING RETURNING COMBAT AIRCREWS, LESSONS LEARNED

Dr. Larry Katz
Army Research Institute, Rotary Wing Research Unit
Ft. Rucker, Alabama 36332

Kurt Kline
Center for Team Performance, Training Development and Implementation Group
Dynamics Research Corporation, Andover, Massachusetts 01810

This paper describes an experiment to evaluate the effectiveness of the U.S. Army's Aircrew Coordination Training Enhancement (ACTE) program delivered by distance learning. A large-scale experiment was designed and executed using three groups of aircrews that received either electronic classroom-based instruction with instructor facilitation on site, distance learning training using the unit's local Digital Training Facility with the primary instructor off site, or no training. Aircrews with varying levels of experience recently returning from combat were evaluated using event-based scenarios performed in the Aviation Combined Arms Tactical Trainer (AVCATT). Measures were developed within Kirkpatrick's (1998) framework. Execution of the experiment was hampered by a variety of factors. One factor was intermittent weather related power outages which made individual crew stations unavailable for short periods of time. This challenge was addressed within a mission contingency framework. A second factor was the participating aircrews' limited experience with the AVCATT trainer which was installed at the aircrew's home station during the units' deployment to combat. Another factor was crew turbulence related to the supporting units' deployment status. Workarounds for administrative and procedural challenges were devised to maintain the integrity of the experiment to the maximum extent possible; however, the evaluation goals of the experiment were not achieved. Results of this experiment are discussed from the perspective of lessons-learned from conducting field research using operational units in wartime.

Introduction

Aircrew Coordination Training (ACT) and Crew/Cockpit Resource Management (CRM) programs were instituted in the 1980's, first in commercial aviation and later in military aviation, to address adverse mishap rate trends that showed the inability of many aviators to work well together in periods of high stress or workload (Helmreich, Merritt, & Wilhelm, 1999). Minor aircraft malfunctions were resulting in fatal accidents with alarming regularity. While aviators generally displayed excellent knowledge and understanding of aircraft systems, operating procedures, rules and regulations and other technical information, they often displayed a glaring inability to communicate effectively, distribute workload, maintain or regain situational awareness and make sound decisions. Military aviation took note of the success of CRM in the civilian sector and instituted similar training programs (Orlady & Foushee, 1987).

ACT/CRM programs have been structured in various ways and continue to evolve as the perspective changes as to what constitutes effective team coordination training. Most programs include the following basic elements:

- A discussion of the core behaviors or basic skill sets that make up ACT. Each program structures these core behaviors differently, but all contain common elements.
- An examination of the applicability of ACT behaviors in the "real world." This typically takes the form of one or more case studies of real-world incidents or accidents and includes an analysis of where or when proper ACT behaviors could have been employed.
- Some type of role-playing or practice of ACT behaviors in a simulated mission setting, i.e., line-oriented flight training (LOFT) or its equivalent.
- Some form of assessment of the learning or changes in attitudes and behaviors that have taken place as a result of the training, and the evaluation of the training by the students.

Research and Development

ARI worked closely with Army aviation training, evaluation, and safety personnel to develop, validate, and field an ACT Exportable Training Package in 1992. Army ACT program performance methods and measures included:

- ACT behaviors or Basic Qualities evaluated with supporting behaviorally anchored rating scales
- Aircrew Training Manual task performance
- Mission performance of two flight simulator scenarios similar in difficulty in terms of time stress, navigational demands, quantity and capabilities of simulated threat.

Initial ACT products developed were validated using traditional pre and post validation methods (Simon, R., & Grubb, G., 1993) and were fielded using a train the trainer system by the US Army during the period 1994-1998 to all active and reserve component aircrews.

Aircrew Coordination Training Enhancement (ACTE)

Commanders and aircrews alike acknowledged the benefit of the mandatory, one-time training support package (Department of the Army, 1992) that was received by all aviators within the Army aviation community. The initial program did not address sustainment issues and did not package the training in a program that would facilitate such training. Lack of effective aircrew coordination continues to be cited as a definite or suspected contributing factor in aviation flight accidents, and it is a factor limiting attainment of the full mission effectiveness of Army aviation. For example, the Director of Army Safety reported in the December 1999 issue of *Flightfax*, "In fact, FY99 produced Army aviation's worst safety performance since Desert Shield/Desert Storm." The ACT program has not been updated since its original introduction. Currently, ACT is conducted in the classroom (Eight hours of instruction with a two-hour, 50 question, multiple-choice exam) with no follow-on mandatory training periods in either aircraft simulators or in the aircraft. Instructors responsible for evaluating and reinforcing this academic training receive four hours of academic training with no exam to determine competency. Temporary measures such as awareness videos, assistance visits, safety newsletter articles, and a web-based training support package have been ineffective substitutes for focused ACT training.

Approach to Revitalize and Sustain Army ACT

The objective of the research effort to enhance Army ACT is to improve the crew and team coordination effectiveness of Army aircrews in their day-to-day mission planning and flight operations. The enhancement program managed by ARI is a multi-year, multiphase program of applied research structured in three major phases – upgrade and sustain the existing ACT program, refresh and maintain the upgraded ACT program, and deploy advanced ACT applications. ARI's Rotary-Wing Aviation Research Unit convened a working group at Fort Rucker to provide guidance and oversight for Army Aircrew Coordination Training Enhancement (ACTE) program. The group is made up of key personnel from the US Army Aviation Center (USAAVNC) and other subject matter experts who serve as contributors to planning, developing, implementing or evaluating the program.

Phase I of the enhancement effort to upgrade and sustain the current ACT program applied the following general approach:

- Analysis of the current aircrew coordination training program from a total systems perspective to identify conflicts, bottlenecks, and deficiencies in implementing team coordination in daily flying operations.
- Refinement of team evaluation techniques and tools for assessing overall performance along specific behavioral proficiency dimensions.
- Development of prototype focused interventions for training and evaluating team coordination behaviors and for managing risk.
- Validation of prototype team training and evaluation techniques in selected aviation units.
- Field-testing of prototype training, evaluation, and technology products.

Phase II of the enhancement program built on the initial research conducted in Phase I and added the necessary courses and data collection events to implement ACT at all levels of aircrews. These courses include:

- Non rated crew member course (NCM) course, the first Army course that recognizes the specific issues of ACT as seen from the mission crew view point
- Core and Advance Aircraft courses, these courses developed especially for the US Army Aviation Center initial entry training supports a building block approach to initial ACT instruction during the 9 month flight school program
- Train the Trainer Course, This course recognized the need for a standardized training and certification program that not only recognized the ACT behaviors and evaluation system but the need to instruct on courseware delivery, facilitation and courseware management to a target audience that has little or no experience in distance learning delivery.
- Delivery of the Train the Trainer program to include collection of end of course survey data.
- Development of the Crew Team Reporting System (CTRS) an anonymous web hosted ACT incident

reporting system to capture data not currently tracked or reported on in Army aviation.
- Pocket Aircrew Guide, this guide was developed and evaluated to assist aircrews in recognizing correct and incorrect behaviors and debriefing missions to facilitate improvement.

In the final phase of ACT program improvement, Phase III, we utilized the guidance in the ACT Master Plan to focus on deploying advanced ACT applications to complete the enhancement program. The desired results of Phase III were to affect the Army's overall aircrew training and evaluation system, risk management and systems safety processes, and daily flight operations in actual aircraft, system simulator, or while conducting training in distributed interactive simulation environments such as the AVCATT or Longbow Crew Training Systems (LCTS).

Collective Research Project

Phase III research was established to deploy advanced ACT applications that focused on:

- Evaluating the effectiveness of ACTE prototype courseware delivered via Distance Learning delivered training. Of particular importance was to address the persistent question of Distance Learning (DL) effectiveness by capitalizing on our database of interactive multimedia courseware delivery via LAN and the demonstrated DL capability of the prototype ACTE courseware. Evaluating learning interaction (e.g., facilitator-learner, learner-learner), adult learning feedback, courseware content control, and testing results reporting issues.
- ACT event-driven scenarios for multiple aircraft missions in advanced simulators and distributed interactive simulation training exercises development. The development of company and battalion level risk management and team coordination methods and measures to address both crewed systems (aircraft) and organizational (C2) leader-focused team training (e.g., collective training scenarios) effectiveness.
- Evaluating the effectiveness of the enhanced ACT program on the operational mission effectiveness and reduction of crew related errors. Conduct a definitive evaluation of the effect of ACT on operational mission effectiveness and reduction of crew related errors. Conduct behaviorally-anchored rating scale (BARS) reliability and validation testing, develop operational mission effectiveness measures and incident reporting procedures to support comparing a unit with enhanced ACT compared to a unit without. Develop ACT event-driven scenarios for multiple aircraft missions in advanced simulators and distributed interactive simulation training exercises.

Developing Evaluation Tools and Techniques

The second task of ACTE Phase I effort was to develop and implement an evaluation methodology for measuring effective performance of aircrew coordination behaviors. The measurement of aircrew coordination behavior is a critical component of the aircrew coordination program and is central to the training content design and delivery. The product of this task is a set of observable measures of individual and collective behavior, the Behaviorally Anchored Rating System (BARS). The BARS provides a readily usable evaluation tool that trainers and ACT facilitators use to teach aircrew members how to apply the BARS as a fundamental means of evaluating aircrew and team performance of ACT behaviors and skills. The vehicle for documenting these evaluations is the ACT Performance Evaluation Checklist which is based on the 5 Crew Coordination Objectives (CCO) and 13 Basic Qualities (BQ) accepted by the Army as descriptors of aircrew coordination behavior. ACT behaviors and skills are organized by CCO and are rated using a seven-point scale with values ranging from 1 (Below Standards) to 7 (Exceeds Standards). Written descriptions are provided for the ACT behaviors and skills and levels of performance for rating aircrews at the values of 1, 4, and 7. These descriptions serve as behavioral "anchors" and are designed to assist in determining how well an aircrew performs ACT behaviors and skills in relation to a well-defined set of performance criteria. The anchors are used as the standard for evaluating ACT performance. This avoids the trap of norm referencing, i.e., comparing one aircrew's performance with that of another. An aircrew's performance is always rated solely in relation to the "anchors." This has long-term implications for the objective measurement of aircrew coordination improvement. (Appendix A)

Once the crew level evaluation tools such as the BARS system was in place the next level to review in the research was the inter and intra team level coordination. As an additional measure the BARS rating system was to be modified using experience from Battle Command Team Training Behaviors (Grubb, Crump et. al. 2001) into a combined battle staff proficiencies measurement system The base was the ACTE BARS and the Battle Staff Performance Evaluation System Check List

(Appendix B) combined into a initial version or V1 for the research event.

Along with the BARS and Battle Master Instruments, scenario event data collection sheets (Appendix C) were created along with the simulator scenarios. Measurements were established to collect data at various points in the research event and include measurements as shown in Figure 1. The Crew Team Reporting System (CTRS) was developed to support follow on incident data collection.

Measurement Area	Measurement Instruments
Demographic Data	• Data Management System Menu Items
Course Critique Questionnaires	• Data Management System Scalar Critique Items • Data Management System Open-ended Items
ACT Knowledge Test	• Data Management System Multiple Choice Items
ATM Task Performance	• Scenario Worksheets
ACT Behaviors	• Performance Evaluation Checklist • Behaviorally Anchored Rating System (BARS)
Mission Effectiveness	• Scenario Worksheets
Risk Management	• Scenario Worksheets
Crew Team Event Reports	• Web Page Menu Items • Web Page Open-ended Items
Battle Staff Survey	• Questionnaire Scalar and Open-ended Items

Figure 1. *Measurements*

Collective Experiment Coordination

Coordination for experiment participants began in October 2003 with the primary focus on Ft. Hood, Texas. The Army's only fully operational collective trainer, the Longbow Crew Training System (LCTS) was in place and operational training up to 6 aircrews at a time in a device that is certified to conduct individual along with collective tasks. Problems with the units available to evaluate at the collective level were:

- Units in varying levels of readiness training, units were only ready for collective level training during a limited time period in the training program.
- Utilization of the LCTS was high, as a one of a kind device the ability to wire in for individual crew monitoring was not a preferred method.
- Unit command structure was hesitant to put additional tasks on the already overloaded schedule of the support staff
- Units in training had already fallen behind Department of the Army mandated dates for unit deployment.

The preceding factors required a new focus on the unit to be selected for this training. It became apparent that a cohesive unit, full trained, not involved in the war on terror and co-located with a suitable collective training device would be difficult to locate.

Aviation Combined Arms Tactical Trainer (AVCATT)

The AVCATT-A system is a dynamic, alternative instructional concept to train and rehearse, through networked simulation, in a collective and combined arms simulated battlefield environment. It supports institutional, organizational, and sustainment training for Active and Reserve Component aviation units worldwide. Collective and combined arms simulation exercises provide commanders with a capability to conduct unit collective training and rehearsals, the unit's mission essential task list and combined arms wartime mission performance requirements. AVCATT-A is a mobile, transportable, tractor trailer based virtual simulation training system designed to provide aviation the capability to conduct realistic, high intensity, task-loaded collective and combined arms training exercises and mission rehearsals.

The physical layout of AVCATT-A consists of two trailers connected by a raised, covered platform. (Figure 2) One trailer includes three reconfigurable manned modules and an 18-person after action review (AAR) facility with an AAR workstation, three dimensional stealth view, plan view (terrain map), and manned module sensor displays. The second trailer includes three reconfigurable manned modules, a battlemaster control room, and a maintenance room. Included in the battlemaster control (BMC) room are the battlemaster console; semi-automated forces (SAF) workstation; unit observer/controller (OC) position; four unit role player (RP) workstations; and overhead stealth, plan view, and manned module sensor displays. Each manned module is reconfigurable to current Army attack, reconnaissance, cargo, and utility aircraft. Each of the four unit RP workstations can be configured as one of six RP functional areas: fire support, ground maneuver, battle command, close air support, logistics, and engineer.

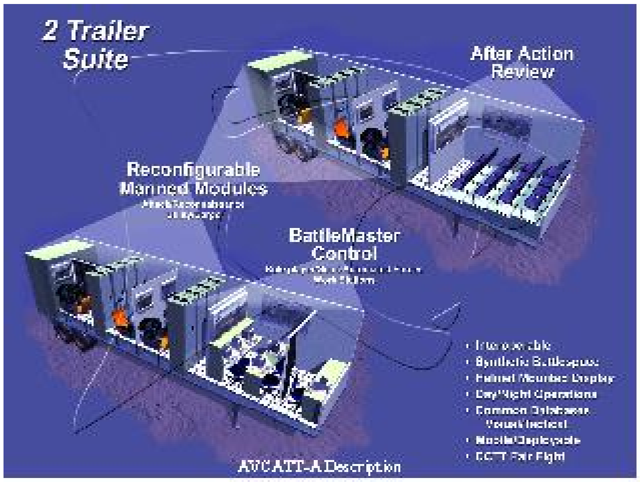


Figure 2. *AVCATT Layout*

The AVCATT was deployed to the final test location in the fall of 2003 and became operational, ready for training in the spring 2004 at the test location. The availability of units returning from the war on terror combined with the operational AVCATT made the selected test location the alternate choice for the conduct of the research data collection.

Available Units

Once the location and collective device was selected the units assigned or in transit back from combat were reviewed. Due to extensive requirements to reacclimatize returning individuals and units and to insure proper reintegration of returning personnel to the local installation time available to conduct the observations of the crews was reduced.

Phase III Collective Experimentation Observer/ Evaluators (OE's) and Battle Master Observers Training

Observer/Evaluator (OE) and Battle Master (BM) training was conducted using US Army Distance Learning Classrooms, contractor instructors presented the Aircrew and Instructor Courses in two-four hour periods. The courses were followed with a training session consisting of the observation of actual crews in simulators followed up by rating using the BARS system to insure inter-rater reliability prior to data collection events. Overall 10 OE/BM were trained and prepared for data collection.

Research Participants

Research participants provided consisted of 12 crews per group for a total of 36 participants. These participants were scheduled for Aircrew ACTE training, pre and post academic testing and pre and post training simulation events.

Lessons Learned

Lesson 1, Ensure all test participants and facilities will be available throughout the course of the entire data collection event.

Test participant availability and status was unknown prior to arrival at the test site. Due to the limited availability of units to participate in research the research director had little room to turn away units who offered to participate. Research participants had been back from overseas combat operations less than 30 days at the start of the research event. During this period research participants were still undergoing mandatory reintegration tasks directed by Department of the Army regulations. During the conduct of the research data collection some participants had mandatory medical appointments and family issues that caused them to miss critical data events. Due to the returning status of many of the installation units, 2 days during the research data collection event the installation was at a minimum manning status, commonly referred to as a "Training Holiday" causing delays in training and support for the research event.

Lesson 2, Crews must be properly trained on simulation devices used to collect research data.

The AVCATT device was a recently fielded system, 12 months prior to research data collection event the Army had begun fielding the system Army wide and less than 90 days at the test location. The device was established as ready for training within 45 days of the beginning of research at the research location. No full scale unit usage of the AVCATT had occurred at the research site prior to this research event. The research participants had not received any training on the AVCATT device and participants needed to be fitted with the helmet mounted visual system and familiarized with both the device and its associated systems just days prior to research data collection. It is important to note that the AVCATT team running the device at the research site worked extended hours and demonstrated professionalism and a get it done attitude that made the best of the situation. Due to the compressed timeline no other training time other than the 1 hour or less familiarization flight was conducted. None of the 36 test participants had any experience on the AVCATT device. During the conduct of the research events it became apparent that the device, designed as a collective trainer and not certified by the Army as an individual task trainer required some level of familiarization not yet determined to develop proficiency in basic flight maneuvers, to include tactical formation flying. Highly experienced, recent combat crews experienced Controlled Flight Into Terrain (CFIT) accidents due to a lack of flying experience in the research device.

Lesson 3, Time must be made available in the research schedule to allow for adaptation of the simulation device for data collection and to address problems in the device operation.

The AVCATT device is designed to conduct collective training and the research required the monitoring of each individual crew by assigned OE's. The design of the AVCATT utilizes helmet mounted displays which precluded the OE's from observing in the actual cockpit, the AAR facilities built into the device will monitor the video from the 5 cockpits but only provides one voice feed. Although satisfactory for the collective training of a unit at the collective level the device would not allow continuous voice monitoring of each crew by their assigned OE. This limitation required the use of an alternate voice monitoring system with a microphone placed inside the crew's headset. This need to place the system into operation and test prior to each event caused delays in an already restricted schedule.

The AVCATT device is a highly complex device sensitive to power fluctuations and computer settings. During the conduct of the test the final scenarios had only been available to the local AVCATT team for 30 days and had not undergone on site testing.

Weather also played a factor. Although the AVCATT is a durable trailer mounted system the fixed power supply is susceptible to lighting and to error on the side of safety two events were delayed due to electrical storm activity in the area. The device was not the only issue; safety of the research participants required the delay. The AVCATT device is placed in an open field with no overhead protection.

Summary

A research based event requiring the participation of over 40 personnel utilizing complex simulation devices with participants conducting multi aircraft missions requires a level of coordination, participant briefings and time delays that cannot be accomplished without a unrestricted timeline, an extensive device familiarization training program and participants focused on the research event at hand.

References

Department of the Army. (1992). Aircrew coordination exportable training package (Vol. 1-3). Fort Rucker, AL: U.S. Army Aviation Center.

Helmreich, R., Merritt, A., & Wilhelm, J. (1999). The evolution of crew resource management training in commercial aviation. The International Journal of Aviation Psychology, 9 (1), 19-32.

Grubb, G., & Crump, S., (2001). Development of a Battle Command Staff Proficiency Measurement System Army Research Institute Technical Report, Washington, DC

Orlady, H. W., & Foushee, H. C. (Eds.) (1987). Cockpit resource management training: Proceedings of NASA/MAC workshop (NASA CP 2455). Moffett Field, CA: NASA Ames Research Center.

Simon, R., & Grubb, G. (1993). Validation of crew coordination training and evaluation methods for army aviation (Tech. Rep. No. E-785U). Wilmington, MA: Dynamics Research Corporation.

ACT Performance Evaluation Checklist For use of this form, see the ACT Aircrew Guide			
CCO	BQ	Crew Coordination Objectives (CCO)/Basic Qualities (BQ)	Rating
1		Establish and Maintain Team Relationships	
	1	Establish and Maintain Team Leadership and Crew Climate	
2		Mission Planning and Rehearsal	
	2	Pre-mission Planning and Rehearsal Accomplished	
	3	Application of Appropriate Decision Making Techniques	
3		Establish and Maintain Workload Levels	
	4	Prioritize Actions and Distribute Workload	
	5	Management of Unexpected Events	
4		Exchange Mission Information	
	6	Statements and Directives Clear, Timely, Relevant, Complete and Verified	
	7	Maintenance of Situational Awareness	
	8	Decisions and Actions Communicated and Acknowledged	
	9	Supporting Information and Actions Sought from Crew	
5		Cross-Monitor Performance	
	10	Crewmembers Actions Mutually Cross-Monitored	
	11	Supporting Information and Actions Offered by Crew	
	12	Advocacy and Assertion Practiced	
	13	Crew/Flight After-Action Reviews Accomplished	
Remarks: (Use continuation sheet[s] if necessary)			
Notes: Consult the ACT Aircrew Guide evaluation procedures and guidelines. Enter a summary rating (1 – 7) in the rating block for each ACT Crew Coordination Objective (CCO). Refer to the rating scale below.			

Below Standards 1	2	3	Meets Standards 4	5	6	Exceeds Standards 7

Appendix A. *ACTE Performance Evaluation Checklist*

Battle Staff Performance Evaluation Checklist			
BSO	BSOF	Battle Staff Objectives (BSO)/ Battle Staff Observational Focus (BSOF)	Rating
1		Develop and Maintain Inter and Intra Team Relationships	
	1	Establish Information and Knowledge Management and Exchange Procedures	
2		Mission Planning, Rehearsal, Roles and Responsibilities	
	2	Decision Authority/Capacity	
	3	Decision Strategies/Manage Debate and Communicate Decisions/Assumptions	
3		Establish and Maintain and Workload Levels	
	4	Prioritize and Select Production Strategies	
	5	Maintain Scanning Across Multiple Decision/Action Items	
4		Exchange Mission Information	
	6	Balance Informational Flow Up and Down Chain	
	7	Maintenance of Battle Space Images and Situational Awareness	
	8	Verify Key Information/Employ Risk Management	
	9	Supporting Information and Actions Sought from Crew	
5		Cross-Monitor Performance	
	10	Anticipate and Prepare for Development of Complex Situations	
	11	Manage Task Priority, Task Sequencing and Information Cost	
	12	Manage Process Error during Staff Rotation and Battle Handover	
	13	Practice Continual Self-Critiques and Lessons Learned	
Remarks: (Use continuation sheet[s] if necessary)			
Notes: Enter a summary rating (1 – 7) in the rating block for each BCO (BSO). Refer to the rating scale below.			

Below Standards 1	2	3	Meets Standards 4	5	6	Exceeds Standards 7

Appendix B. *Battle Staff Performance Evaluation Checklist*

AH-64A AVCATT Scenario Worksheet | Area Recon Scenario 2
PRE / POST Training Mission | Battle Master Observer/OE ______________
Date ________ Period _____ MM # ___________ | Crew # _________________

EVENT: Engage Targets with Indirect Fires

Event Type: Crew Team External	**Event Time:**	**Observed: Y N**

Event Trigger: Team engages targets with indirect fires (SEAD).	Action: Individual crews engage assigned targets with indirect fires in SEAD.

ATM Tasks:	**Observed?**	**Rating:**	**Related Behaviors:**
1079 Perform Radio Communications	Y N	S+ S S- U	CCO1 CCO2 CCO3 CCO4 CCO5
2007 Perform Aerial Observation	Y N	S+ S S- U	CCO1 CCO2 CCO3 CCO4 CCO5
2020 Call for and Adjust Indirect Fire	Y N	S+ S S- U	CCO1 CCO2 CCO3 CCO4 CCO5
2049 Search for and ID Targets with TADS	Y N	S+ S S- U	CCO1 CCO2 CCO3 CCO4 CCO5
2091 Transmit a Tactical Report	Y N	S+ S S- U	CCO1 CCO2 CCO3 CCO4 CCO5

Performance Measures:

Did the team request indirect fire support? Yes / No (Circle one) # Requests ____

Was indirect fire effective (e.g., targets suppressed or destroyed)? Yes / No (Circle one) # Engagements ____
Targets Destroyed _____

Did the team members share indirect fires target engagement information with other team members? Yes / No (Circle one)

Notes:

Overall Event Rating S+ S S- U CCO1 CCO2 CCO3 CCO4 CCO5 (OPTIONAL)

Appendix C. *Scenario Worksheet*

THE AIR TRAFFIC SELECTION AND TRAINING BATTERY: WHAT IT IS AND ISN'T (AND HOW IT HAS CHANGED AND HASN'T)

Raymond E. King
Andrew R. Dattel
Civil Aerospace Medical Institute
Oklahoma City, OK

The Federal Aviation Administration (FAA) has developed a new selection procedure, the Air Traffic Selection and Training (AT-SAT) computerized test battery, to help select Air Traffic Control Specialists. AT-SAT is an aptitude test and not a test of air traffic control knowledge. Of the 264 applicants who have taken AT-SAT, 155 responded to a job announcement, while 109 previously passed the OPM (pre-employment) test and had to achieve a passing score on AT-SAT before they were admitted into training at the FAA Academy. Of the 155 job-announcement applicants, 131 (84.52%) achieved a passing score of 70 or greater (termed a "qualifying score"), while 24 applicants (15.48%) failed to achieve a minimum score of 70. Those who had been prescreened with the OPM test fared a bit better, with 104 (95.41%) achieving a qualifying score; five (4.59%) applicants failed. Current research efforts include equating a parallel form, rehosted on a Windows 2000 operating platform, with the assistance of research participants from the US Army, Navy, and Air Force. Another recent project was focused on reweighting the subtests and adjusting the overall constant to address issues of potential adverse impact, without compromising validity. A greater concern in this effort was to ensure that AT-SAT performance would predict job performance rather than just success or failure in training. Despite this reweighting effort and updating of the operating platform, the content of the battery remains unchanged. Future efforts will involve a longitudinal validation to compare performance on AT-SAT with success in training and on the job.

The development and validation of selection instruments for occupations where a sizeable number of applicants are needed to fill demanding positions play a critical role in reducing costs associated with attrition from training programs. Validation also ensures that those who are hired have (or are likely to develop) the necessary knowledge, skills, and abilities to perform successfully on the job. The duties of an Air Traffic Control Specialist (ATCS), specifically those providing *separation* services, which makes these employees individually responsible for more lives than the practitioners of any other occupation in the United States (Biggs, 1979). The FAA developed the AT-SAT battery to replace a two-stage selection process in which ATCS applicants completed an Office of Personnel Management (OPM) test battery and a nine-week screening program at the FAA Academy in Oklahoma City, OK. This previous selection process proved to be expensive and inefficient (Ramos, 2001). AT-SAT was developed based on the results of the Separation and Control Hiring Assessment (SACHA; Nickles, Bobko, Blair, Sands, & Tartak, 1995) job analysis of the duties of the ATCS options.[1]

[1] There are three options in the 2152 occupational series: terminal, en route, and flight service station. Terminal controllers can be divided into two groups: tower cab and TRACON. AT-SAT is not used for the selection of flight service station personnel.

The SACHA job analysis reviewed the existing ATCS job analysis literature. An extensive assortment of documents was examined for terms suitable to the knowledge database, including FAA, military, and ATCS civilian courses. After reviewing and summarizing the existing job analysis information, the SACHA project staff visited sites to observe controllers from the two options and assignments. Subject-matter experts (SMEs) were also questioned about the qualities they considered necessary for effective job performance. The worker requirements determined necessary for the job of ATCS were then used to design a series of self-administering computerized tests to assess the ability of applicants to perform these tasks.

This paper focuses on the current status and future plans for the recently[2] implemented AT-SAT battery. AT-SAT is a computerized test battery comprised of eight subtests based on 22 individual scores that, when weighted (forming "part scores") and combined, are totaled (with an overall constant added) for an overall score. AT-SAT comprises the following subtests: *Air Traffic Scenarios Test, ATST; Analogies, AY*; *Angles, AN*; *Applied Math, AM*; *Dials, DI*; *Experiences Questionnaire, EQ*; *Letter Factory,*

[2] AT-SAT was approved as the official ATCS selection test, for those applicants without previous air traffic control experience, on May 13, 2002, with June 2002 marking the first time the test was operationally used.

LF; and Scan, SC. AT-SAT is an aptitude test and not a test of air traffic control knowledge. The goal of AT-SAT is to gauge the likelihood of success in air traffic control training and, more importantly, subsequently on the job. Seven of the eight subtests assess aspects of cognitive ability, while one, *EQ*, assesses issues in the personal history/personality realm. Four (*ATST*, *AY*, *LF*, *SC*) of the subtests are dynamic; they are interactive and can only be administered via computer. The remaining four are static, similar to pencil-and-paper tests, but are administered via computer in AT-SAT.

Before operational use of AT-SAT was approved for hiring purposes, FAA employees who were members of minority groups raised concerns over potential adverse impact.[3] Consequently, FAA management met with representatives from the groups to hear their concerns. The concern about the potential for adverse impact against African Americans seemed well founded, as only three out of every 100 black applicants were predicted to achieve a score of at least 70 (the minimum passing score – termed a "qualifying score") on AT-SAT. The issue went beyond pass rates of minority applicants. By design, 38% of fully certified incumbent FAA controllers would not pass AT-SAT under the original scoring scheme. The original passing score of 70 had been calibrated so that only 62% of incumbent fully certified controllers would achieve an AT-SAT score equal to, or greater than, 70 in an effort to minimize FAA Academy failures and to compensate for the need for ATCSs to perform potentially more difficult duties in the future. The goal was to at least preserve and strive to improve the level of functioning in the workforce (Waugh, 2001).

In response, the FAA requested that scientists review the weights of AT-SAT subtests to reduce adverse impact. At the same time, there was an emphasis on maintaining the overall validity of the battery. Additionally and more importantly, management made the case that the cut score should be set at the point where most fully qualified incumbent FAA controllers would pass FAA's entry-level aptitude test. Consequently, the AT-SAT subtests were re-weighted and the constant was adjusted. The content of the subtests themselves was not changed, rather the subtests were weighted differently. The challenge was to retain adequate validity while reducing adverse impact. Test validity (job-relatedness) is determined by the strength of the correlation between the test score and job performance measures. After reweighting, the correlation between AT-SAT and job performance was reduced slightly, from .69 to .60. Compared with most validation coefficients, this is still a strong relationship with job performance. The relationship with job performance is especially important in this context as any remaining adverse impact can be justified by business necessity. In the end, however, it was found that potential adverse impact for women and Hispanics had been completely eliminated and had been greatly reduced for African Americans. Adverse impact will be continually assessed with job applicants. Finally, to further address the potential problem of adverse impact, the FAA decided to abandon a strict "top-down" approach to hiring and instead use a category ranking method. Under this scheme, successful examinees are divided into two groups: those scoring 85 and above (termed "well qualified") and those scoring from 70 to 84.9 (termed "qualified"). Those in the well-qualified group will be offered employment before anyone in the qualified group. Within the ranges, veterans are hired before non-veterans, but selecting officials can consider other job-related factors, such as the ability to speak English well enough to be understood and self-reported interest in the job, dimensions that are not measured by AT-SAT.

A forthcoming study (Dattel & King, in preparation) applied the weights and additive constant developed to address potential adverse impact to the scores of 292 voluntary research participants who took the AT-SAT under the original scoring scheme. This rescoring increased the research participants' overall scores by an average of 9.08 points, with the scores of Caucasians increasing by 8.84 points, African Americans by 9.82 points, and Hispanics by 11.03 points. Additionally, this rescoring increased the overall pass rate (scores equal to or greater than 70) in this sample from 36.3% to 68.2%. It is important to bear in mind that these test takers were not applicants and were, instead, voluntary research participants.

It should be noted that there are several applicant categories whose members do not have to take and pass AT-SAT to be considered for employment. Military controllers and Department of Defense civilian controllers are included in this category as well as former PATCO controllers who are now eligible for rehire. These applicants still face a competitive process and are by no means

[3] Adverse Impact – "A selection rate for any race, sex, or ethnic group which is less than 4/5 (80%) of the rate for the group with the highest rate" Uniform Guidelines on Employee Selection Procedures (1978), Sec 4D.

automatically hired; they are just exempt from having to take AT-SAT.

The report, *A Plan for the Future: The FAA's 10-Year Strategy for the Air Traffic Control Workforce (http://www.faa.gov/newsroom/controller_staffing/WorkforcePlan.pdf),* was submitted to the U.S. Congress in December 2004. This report provides a plan to mitigate pending controller retirements and contemplates strategies to achieve appropriate staffing levels. While previously military applicants with air traffic control experience were able to satisfy many of the FAA's hiring needs, there is a need to begin hiring more controllers. The availability of applicants with previous experience will quickly be exhausted. AT-SAT will thus become an instrument of increasing importance. How did the hiring need become so urgent? An overwhelming majority of the air traffic control workforce went on strike on August 3, 1981. During this time, President Ronald Reagan ordered the striking controllers to return to duty within 48 hours. When 10,438 (out of a workforce of approximately 15,000) striking controllers did not return to work in this timeframe, the president fired them. Facing a sudden shortage of controllers, the FAA hired 3,416 individuals in 1982 and another 1,720 in 1983. From 1982 through 1991, the FAA hired an average of 1,527 individuals per year. The majority of entrants met the 18 to 30 years-of-age entry requirement. This hiring wave created the potential for a large portion of the controller workforce to reach retirement age at roughly the same time. Based on recent projections, over the next 10 years, 73 percent of the agency's 15,000 controllers will become eligible to retire. Total losses over the next 10 years are expected to be nearly 11,000 (FAA, 2005).

The Current State of Affairs

To date, 264 applicants have taken AT-SAT as part of their job application process; 155 of these applicants responded to a job fair announcement (soliciting applicants for a specific position), while 109 had previously passed the OPM test (pre-employment test) and had to achieve a passing score on AT-SAT before they were admitted into training at the FAA Academy. Of the 155 job fair applicants, 131 (84.52%) achieved a score of passing score of 70 or greater, while 24 applicants (15.48%) failed to achieve a minimum score of 70. Those who had been prescreened with the OPM test fared a bit better, with 104 (95.41%) achieving a qualifying score; five (4.59%) applicants failed. AT-SAT was also taken by 727 research participants. These participants were students enrolled at the Academy but took the AT-SAT voluntarily (their enrollment was obtained via voluntary consent and their continued employment was not contingent on their performance on AT-SAT). This group includes, but is not limited to, retired military personnel and graduates of collegiate training initiatives (CTI) who were previously hired with the OPM test.

Figure 1 presents overall AT-SAT results in a continuous, as opposed to a dichotomous (pass/fail), fashion. To aid in the comparison of results between groups (job announcement or "job fair" applicant, OPM applicant, research participants), all results have been transformed into the current weighting scheme. The groups are significantly different, ($F(2,930) = 38.440$, $p<.001$), with OPM applicants significantly outperforming Job Fair applicants and Research participants. Job Fair applicants significantly outperformed research participants.

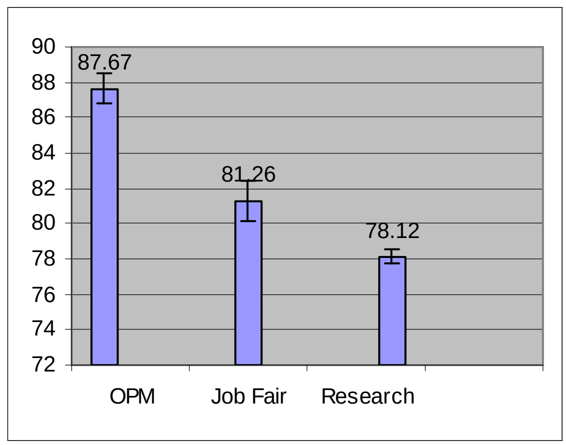


Figure 1. *Overall AT-SAT Score by Participant Category (weighted under current scheme). Error bars indicate standard error.*

The best way to appreciate results of AT-SAT is to consider the battery subtest by subtest (see Figure 2). As in Figure 1, all subtest scores have been transformed using the current weighting scheme. However, the means of the sub-test scores have been converted to standardized scores (z-scores[4]) for a more consistent presentation. There are significant group differences on six of the eight subtests (no significant differences were found between groups for ATST or AY), as delineated in Table 1.

[4] Z-scores range from a high of 1 to a low of –1 with 0 as the mean.

Sub Test	Means (SD)	Post hoc group differences (∝=.05)
DI *F*(2,930)=13.24, *p*<.001, *Mse*=.048	OPM=1.91 (.19) Job Fair=1.85 (.20) Research=1.79 (.22)	OPM > Job Fair OPM > Research Job Fair > Research
AM *F*(2,930)=19.08, *p*<.001, *Mse*=34.901	OPM=21.00 (5.34) Job Fair=16.18 (6.57) Research=19.40 (5.88)	OPM > Job Fair OPM > Research Research > Job Fair
SC *F*(2,930)=53.77, *p*<.001, *Mse*=4.091	OPM=9.77 (.77) Job Fair=9.33 (1.45) Research=7.90 (2.20)	OPM > Research Job Fair > Research
AN *F*(2,930)=14.41, *p*<.001, *Mse*=.060	OPM=1.67 (.20) Job Fair=1.53 (.26) Research =1.53 (.25)	OPM > Job Fair OPM > Research
LF *F*(2,930)=80.86, *p*<.001, *Mse*=4.282	OPM=6.30 (1.47) Job Fair=6.29 (1.50) Research=4.21 (2.21)	OPM > Research Job Fair > Research
ATST *F*(2,930)=3.01, *p*=.050, *Mse*=.354	OPM=2.08 (.52) Job Fair=1.98 (.59) Research=1.93 (.60)	
AY *F*(2,930)=1.92, *p*=.147, *Mse*=1.881	OPM=5.61 (1.21) Job Fair=5.26 (1.32) Research=5.48 (1.40)	
EQ *F*(2,930)=105.02, *p*<.001, *Mse*=49.773	OPM=33.50(4.52) Job Fair=32.72 (5.76) Research=25.01 (7.49)	OPM > Research Job Fair > Research

Table 1. *Group Differences, Subtest by Subtest.*

While the superior performance of the OPM applicants is not surprising, given that they were previously screened with the OPM test, one should view the weaker performance of the Research participants with a degree of caution. These participants, even though some were also pre-screened with the OPM test, may have been less motivated to do their very best as they had already been hired and were explicitly told that their performance on AT-SAT would not impact their employment with the FAA.

What's Current and What's Next?

Only one version of the AT-SAT battery was constructed during the initial development and validation effort, meaning that all persons who took AT-SAT received the same items and in the same order. Consequently, there was an increased likelihood that any improvement in the score of someone who retook the test was due to a practice effect (Heil, Detwiler, Agen, Williams, Agnew, & King, 2002). The use of one version (or "form") also suggests that the test may be more vulnerable to coaching since there is only one set of items that must be trained. The result is a potentially incorrect hiring (false positive) decision, with an increased likelihood that such an applicant would not be ultimately successful. A score inflated as a result of coaching does not increase the individual's actual ability to perform air traffic control work. To guard against the empirically demonstrated compromising effects of practice and coaching (Heil et al., 2002) and to mitigate against the deleterious results of the security of AT-SAT being compromised, an alternate version has been developed. The first step in this process, the "pilot study," was to develop alternative items and test them on volunteer research participants to ensure that they were at the appropriate level of difficulty. The U.S. Air Force and Navy graciously supplied these participants from air traffic control schools at Keesler Air Force Base, MS, and Pensacola Naval Air Station, FL, respectively. The end result was two parallel forms. Current research involves equating these parallel forms, rehosted on a Windows 2000 operating platform, an upgrade from Windows 95. The U.S. Army at Fort Rucker, AL, has joined its sister services in collaborating in this effort by supplying participants who are either air traffic controllers or students in air traffic control training. For adequate statistical power, the goal is to collect data from a total of 1,500 participants across these military sites. Each military participant completes two four-hour test sessions in the course of a day. While these research participants may differ from actual job applicants, they are encouraged to do their very best. When completed, the parallel version of AT-SAT will be comprised of the same subtests, with similar items. The tests will be presented in a standardized fashion.

Currently, AT-SAT is not used for placement decisions; that is, scores are not used to assign successful applicants to centers or terminal facilities. There is growing interest in determining if AT-SAT can be effective in placing new hires into facilities. Other future research efforts include longitudinal validation: comparing performance on AT-SAT with success in training and on the job. The ultimate goal of research with AT-SAT is to ensure that those selected to enter the ATCS career field possess (or will develop) the necessary knowledge, skills, and abilities to ensure that air traffic moves in a safe and expeditious manner.

References

Biggs, D. (1979) *Pressure cooker.* New York: Norton.

Dattel, A. R. & King, R. E. (in preparation). Gauging the impact of attempts to reduce potential adverse impact.

FAA (2005). *FAA announces 10-year air traffic controller staffing plan*. Retrieved January 20, 2005 from http://www.faa.gov/newsroom/controloler_staffing/WorkforcePlan.pdf. Federal Aviation Administration.

Heil, M. C., Detwiler, C. A., Agen, R. A., Williams, C. A., Agnew, B. O., & King, R. E. (2002). *The effects of practice and coaching on the Air Traffic Selection and Training Battery* (DOT/FAA/AM-02/24). Washington, DC: FAA Office of Aerospace Medicine.

Nickels, B. J., Bobko, P., Blair, M. D., Sands, W. A., & Tartak, E. L. (1995). *Separation and control assessment (SACHA) final job analysis report*. Bethesda, MD: University Research Corporation.

Ramos, R. A. (2001). Air traffic selection and training (AT-SAT) project. In R. A. Ramos, M. C. Heil, & C. A. Manning (Eds.) *Documentation of validity for the AT-SAT computerized test battery, Volume 1* (DOT/FAA/AM-01/5) (pp. 1-5). Washington, DC: FAA Office of Aerospace Medicine.

Waugh, G. (2001). Predictor-criterion analyses. In R. A. Ramos, M. C. Heil, & C. A. Manning (Eds.) *Documentation of validity for the AT-SAT computerized test battery, Volume II* (DOT/FAA/AM-01/6) (pp.37-42). Washington, DC: FAA Office of Aerospace Medicine.

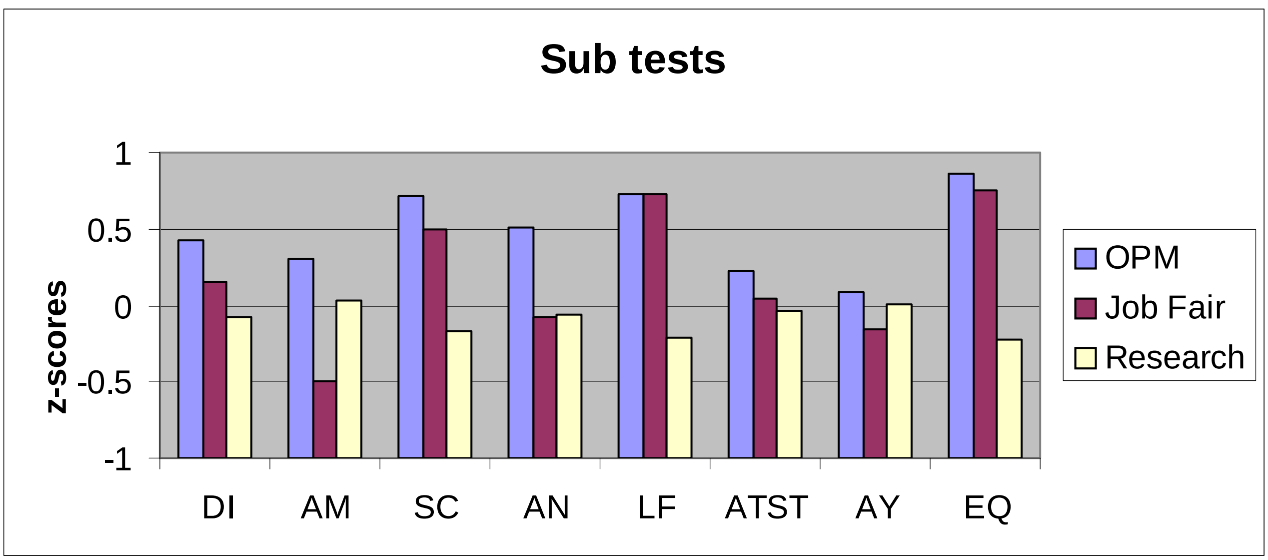


Figure 2. *Group Differences, Subtest by Subtests*

GENERAL AVIATION VFR-INTO-IMC: *Z*-SCORE FILTERING OF DEMOGRAPHIC AND PERSONALITY VARIABLES, AND THE PERSONALITY PARADOX

William Knecht
Civil Aerospace Medical Institute, FAA
Oklahoma City, OK

Does pilot personality affect risk-taking with weather? Armchair logic says "Yes," while data often say "No." In this work, we apply the technique of *z*-score filtering (slice analysis) to pilot takeoff decisions made in the face of simulated adverse weather seen at taxiway level. Such a filtering technique might prove useful, provided emphasis is kept to maintain experiment-wise reliability. Statistical and methodological problems with personality data are discussed. The results of this particular data set showed a strong effect of weather on takeoffs, as measured by visibility, cloud ceiling, and the interaction of the two. But, despite best efforts, no strong effect of personality could be found in this data set. Theoretical reasons are discussed as to why it may be difficult to show that personality predicts behavior.

Introduction

Visual flight rules-flight into instrument meteorological conditions (VFR-into-IMC) is a serious problem in general aviation (Adams, Koonce, & Hwoschinsky, 2002; Hunter, 2002a,b; O'Hare, 1990; O'Hare & Owen, 2002; O'Hare, Chalmers, & Scuffham, 2003; Wiegmann, Goh, & O'Hare, 2002). The FAA has identified VFR-into-IMC as a leading cause of GA fatalities, and has made it a top priority in its 2004 and 2005 *Flight Plan* (FAA, 2004).

It is natural to wonder if pilots' personality influences their risk for venturing into severe weather. Armchair logic says "Of course it does." However, personality tests have a mixed record for being able to predict behavior. This has been called "The Personality Paradox"—the notion that, somehow, personality *must* exist and *must* affect behavior—yet the connection is usually hard to demonstrate.

In aviation psychology, at least one author asserts that virtually all personality research on pilots can be shown to have at least one fatal flaw (Besco, 1994). Besco cites a host of methodological errors, such as weak validation procedures, lack of replication, experimenter biases, "potential for fakery" of responses, and lack of objective performance criteria. Any of these flaws renders research results suspect.

There are also theoretical reasons why personality tests may not predict behavior. Within the field of personality research, a great "Person-Situation Debate" has raged for years. A good summary of this is given in Epstein & O'Brien (1985). To sum up briefly, every behavior is probably specific to some rather narrow environmental context, or *domain* (Weber, Blais, & Betz, 2002). For example, roads and skies are two different domains. A risky driver may not necessarily be a risky pilot. This means that domain-specific tests normed in a non-aviation domain may not have much application to aviation.

A central theoretical issue here is whether or not there even exist any such things as "domain-free personality traits." Such traits would have to be stable and exert an influence on behavior, no matter in what context that behavior took place.

In the present work, a number of common personality measures were examined, as well as two demographic factors commonly assumed to correlate with risk-taking behavior (pilot age and number of flight hours). The idea was to see if any of their scores, or sub-scores could predict takeoff into adverse weather.

Method

Thirty general aviation (GA) pilots were first given an extensive battery of common personality tests (Table 1). Pilots were next positioned on a taxiway in a flight simulator and were told that their aircraft was not currently certified for instrument flight, so any takeoff would have to be VFR. Three levels of simulated ground visibility (V = 1, 3, 5 statute miles) and two levels of cloud ceiling (C = 1000', 2000') were manipulated as independent variables in a 3x2 between-subjects design. Each pilot saw one V,C combination and then had to decide whether or not to take off and fly in that weather. Logistic regression modeling was then conducted to see if personality test scores could predict actual yes/no takeoff decisions.

Instrument	High score implies	Reference
Aviation Safety Attitude Scale	high history of aviation risk behavior	Hunter, 1995, 2002a, 2002b
Anxiety Sensitivity Index	high scores indicate high anxiety	Peterson & Reiss, 1994
Barratt Impusiveness Scale V10	high impulsivity	Barratt, 1975
Eysenck Impulsivity Scale	high impulsivity	Eysenck & Eysenck, 1964,1985
Hazardous Events Index	high history of aviation risk behavior	Hunter, 2002b
Multidimensional Personality Questionnaire	high degree of specified trait	Patrick, Curtin, & Tellegen, 2002
Risk Orientation Questionnaire	high risk tolerance	Rohrmann, 2002
Sensation Seeking Scale	high desire for stimulus-seeking	Zuckerman, 1994
State-Trait Anxiety Scale	high anxiety	Spielberger, 1983

Table 1. *List of personality tests examined in this study.*

Results

Predictably, the single most significant groupwise factor in pilots' decisions turned out to be the weather itself. Seventy percent of pilots chose to stay on the ground. Contrast this with an expected rate of 100% takeoffs, had there been unlimited visibility and ceiling ($p<<.0001$ by binomial expansion, assuming a highly conservative 28/30 takeoff ratio).

Throughout the regression analysis, despite extensive attempts to predict takeoff through seemingly sensible combinations of demographic and personality factors, no model ever seemed to explain much more outcome variance than did weather all by itself (about 50%).

Was this to say that pilot personality did not matter? Or was it was more likely that each pilot had a unique, individual set of motivations and propensities—a "story," if you will—but that there were so many individuals with so many different stories that it made groupwise analysis difficult?

To try to get at these individual stories, predictor scores were converted to *z*-scores, and then threshold-filtered to try to reveal patterns of predictors whose absolute values were high relative to the group mean.

This kind of slice analysis has potential as an analytical technique, particularly in cases where we wish to tell stories about a relatively small number of individuals. However, we do need to keep in mind the effect that looking at many predictors will have on experiment-wide (familywise) error (Keppel, 1982).

These potentials and issues are best seen through example. In this experiment, two pilots chose to take off into the very worst weather presented (1 mile ground visibility plus 1000' cloud ceiling). What, if anything, set these two pilots apart from the other 28?

Using slice analysis, an initial *z*-threshold value (θ_z) of 3.3 standard deviations was established. This theta value corresponded to the Bonferroni correction necessary to maintain familywise error at $\alpha = .05$ (two-tailed), despite the examination of 28 predictors for two subjects. The corrected α was derived from the desired familywise α divided by the number of examinations planned (.05 / (2*28) = .0009. The *z*-level necessary to achieve that new α was $z_{critical} = \pm 3.3$, which yielded an area of .0009 under both tails combined. This can be cross-checked by expanding the binomial $(\alpha, 1-\alpha)^n$ for n=56 factors and noting that $(1-.0009)^{56} = .95 = 1-\alpha$, which equals the chance of zero Type 1 errors (the chance of finding no statistical "significances" where none truly exist).

As Figure 1 illustrates, $\theta_z = 3.3$ was a very stringent criterion. All that remained after thresholding was a single surviving predictor for a single pilot (the first two variable slots merely represented visibility, and ceiling, which were not thresholded). This surviving predictor was the Hazardous Events Index (HEI) score, which measured pilots' past history of hazardous encounters. So it did make sense that an elevated HEI score could relate to risk-taking in this scenario.

At this point, it made some sense to try relaxing the familywise α to assess how this would trade off in terms of increased information. Relaxing to $\alpha = .10$ gave a $\theta_z = 3.1$. That still left 90% assurance that the overall analysis was reliable, which still translated to a best guess of zero expected overall Type I errors. This produced at least one extra piece of information about S 2031, as Figure 2 shows.

Unfortunately, the surviving predictor was a below-average Rohrmann Risk Orientation Questionnaire, *Risk Propensity* index score (ROQ-P). Having a *low* propensity for risk was inconsistent with this pilot's actual takeoff into the very worst conditions. So that left a logical quandary.

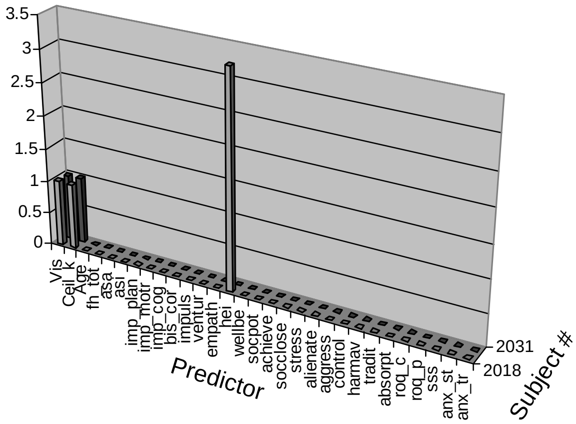


Figure 1. *Predictors thresholded at $\theta_z = 3.3$. This criterion was so rigorous that it failed to show anything other than an elevated Hazardous Events Index (HEI) score for subject 2018.*

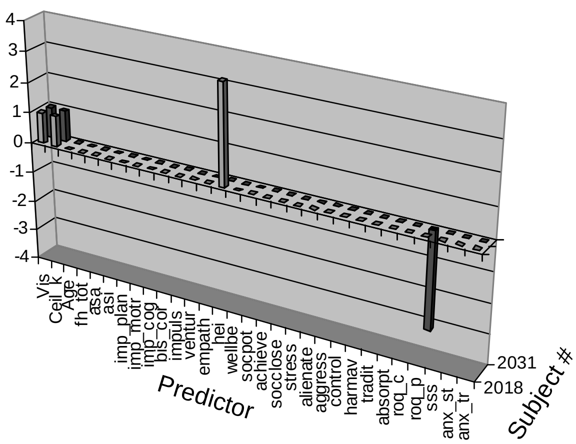


Figure 2. *$\theta_z = 3.11$. Familywise reliability is .90, expected Type I errors still = 0.*

To carry this filtering technique to its conclusion, θ_z was finally lowered all the way to 1.5. This provided only slightly more information, and led familywise reliability to plunge to .0003, with seven expected Type I errors, despite only four predictors surviving threshold.

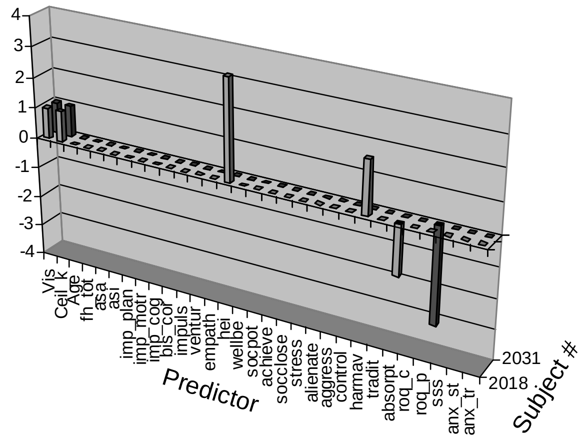


Figure 3. $\theta_z = 1.5$. *Little information is gained, despite a great loss in reliability.*

Discussion

The implication this methodology has for analysis is mixed. On the one hand, it makes it quite easy to imagine a "story" for each pilot—some pattern of predictor scores that might explain why that pilot acted as he or she did in some circumstance. On the other hand, elements of these stories may not be reliable or even make logical sense. In fact, as we can see with this data set, by the time we lower our reliability threshold (θ_z) to a level where we can see emerging patterns, our familywise error rate is in trouble. That means that, under certain circumstances, we could have gotten strong-looking—but counterfeit—patterns simply from random numbers.

So does this mean that pilot personality had nothing to do with pilot behavior? Not necessarily. What seemed more likely was that:

1. Aviation-specific versions of most of these predictors may be needed.
2. Even if the right predictors were tracked, scores may not have differed greatly enough from the mean to statistically distinguish themselves from noise. However, their concomitant traits might still have exerted influence on behavior.
3. Combinations of traits may have acted synergistically to create a "whole greater than the sum of the parts."

Point 1 concerns the notion that risk can be domain- and situation-specific. If so, then we would need aviation-specific personality tests, normed on pilots and specific aviation behaviors (e.g. the HEI).

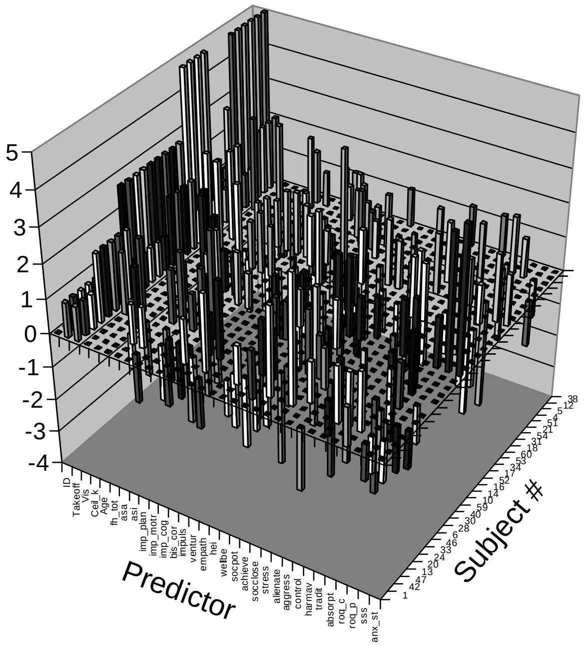


Figure 4. *θ_z = 1. Dense patterns of information emerge. These may have small amplitude, but might still exert true effect on takeoff.*

Figure 4 speaks to Point 2. As this illustrates, by greatly relaxing θ_z we can visualize how each subject may very well have a unique personality profile. But these patterns do seem almost all over the map. The problem is one of reliably demonstrating patterns when most of them lie "submerged" below a statistical threshold elevated by the number of factors being examined.

Points 2 and 3, if true, would make the study of pilot personality very difficult, if not impossible. We could call all this part of the mathematical basis for the Personality Paradox.

First, we have a theoretical situation loosely analogous to Heisenberg's Uncertainty Principle (the impossibility of simultaneously knowing a particle's momentum and position). The act of looking for meaningful patterns—examining many factors simultaneously—decreases the statistical reliability of each score to the point where the information becomes untrustworthy. We seemingly cannot have our cake and eat it too.

Second, and equally bad, if synergy does exist between variables, then the situation worsens because of a possible combinatorial explosion. Equation 1 shows the formula for *n* objects taken *k* at a time:

$$\binom{p}{k} = \frac{p!}{k!\left(p-k\right)!} \qquad (1)$$

So, if our personality test has, say, 11 factors, then there are 55 ways we could make pairs, 165 ways for triplets, 330 for quadruplets, and so forth. If, truly, "the action is in the interaction," then, given these kinds of numbers, we run headlong into impossibly strict criteria for limiting familywise error.

In short, we may be statistically caught between a rock and a hard place. The Personality Paradox may be an inevitable mathematical consequence of combinatorics.

Conclusions

It is difficult to dismiss the intuitive notion that "right stuff" personality plays a major role in pilot decision making. A logical next step in pursuing this issue might be to use a "Big Five" OCEAN approach. This would involve testing five commonly accepted factors of Trait Theory: openness to experience, conscientiousness, extroversion, agreeableness, and neuroticism. Popkins (2004) gives an excellent critical review of this approach. Since five is not a very large number, this would go a long way toward reducing combinatorial effects.

Yet we are faced with a burgeoning suspicion that it may be difficult to identify most of the personality factors that putatively affect behavior, even after the fact, let alone before it. There do appear to be so many individuals with so many different "stories" that mathematical arguments arise that indicate groupwise analysis and behavioral prediction on the basis of personality will always be difficult.

References

Adams, R. J., Koonce, J. M., & Hwoschinsky, P. V. (2002). Decision-making styles associated with accidents: Defining the high-risk pilot. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, 2002, 948-52.

Barratt, E. S. (1975). *Barratt's Impulsiveness Scale*. Princeton, NJ: Educational Testing Service

Besco, R.O. (1994, January). Pilot personality testing and the emperor's new clothes. *Ergonomics in Design*, pp.24-29.

Epstein, S., & O'Brien, E.J. (1985). The person-situation debate in historical and current perspective. *Psychological Bulletin*, 98(3), 513-537.

Eysenck, H.J., & Eysenck, S.B.G. (1964). *Eysenck Personality Inventory*. San Diego, CA: Educational and Industrial Testing Service.

Eysenck, H.J., & Eysenck, M.W. (1985). *Personality and individual differences: A natural science approach.* New York: Plenum.

Federal Aviation Administration, (2004). *Flight Plan*. Retrieved December 29, 2004 from http://www.faa.gov/AboutFAA/Plans.cfm

Hunter, D.R. (2002a, September). *Measuring the safety-related attitudes of pilots.* Paper presented at the Twenty-fifth Conference of the European Association for Aviation Psychology. Warsaw, Poland.

Hunter, D.R. (2002b). *Risk perception and risk tolerance in aircraft pilots*, (Report no. DOT/FAA/AM-02/17), Washington, DC: Office of Aerospace Medicine.

Hunter, D.R. (1995). *Airman research questionnaire: Methodology and overall results.* (Report DOT/FAA/AM-95/27). Washington, DC: Federal Aviation Administration.

Keppel, G. (1982). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice-Hall.

O'Hare, D. (1990). Pilots' perception of risks and hazards in general aviation. *Aviation, Space, and Environmental Medicine,* 61(7), 599-603.

O'Hare, D., & Owen, D. (2002). Cross-country VFR crashes: Pilot and contextual factors. *Aviation, Space, and Environmental Medicine, 73*(4), 363-66.

O'Hare, D., Chalmers, D., & Scuffham, P. (2003). Case-control study of risk factors for fatal and non-fatal injury in crashes of civil aircraft. *Aviation, Space, and Environmental Medicine*, 74(10), 1061-66.

Patrick, C.J., Curtin, J.J., & Tellegen, A. (2002). Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment*. *14*(2), 150-63.

Peterson R.A, & Reiss S. (1994). *Anxiety Sensitivity Index Manual.* (2nd Ed.). Worthington, OH: International Diagnostic Systems.

Popkins, N.C. (2004). *The Five-Factor Model: Emergence of a taxonomic model for personality psychology*. Retrieved January 18, 2005 from http://www.personalityresearch.org/papers/popkins.html. Unpublished manuscript, Northwestern University.

Rohrmann, B. (2002). *Risk attitude scales: Concepts and questionnaires. Project report.* Retrieved January 3, 2005 from http://www.rohrmannresearch.net/rohrmann-racreport.pdf. Unpublished manuscript, University of Melbourne, Australia.

Spielberger, C. (1983) State-Trait Anxiety Inventory (Form Y). Palo Alto, California: Mind Garden.

Timberlake, W. (2002). Niche-related learning in laboratory paradigms: the case of maze behavior in Norway rats. *Behavioural Brain Research*, 134, 355-374.

Weber, E.U., Blais, A.-R., & Betz, N. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, 15, 263-290.

Wiegmann, D.A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules flight into adverse weather. *Human Factors, 44*(2), 1890-97.

Zuckerman, M. (1994). *Behavioural expressions and biosocial bases of Sensation Seeking.* Cambridge: Cambridge University Press

SURPRISE AND UNEXPECTEDNESS IN FLYING: FACTORS AND FEATURES

Janeen Kochan
Eyal Breiter
Florian Jentsch
Team Performance Laboratory
University of Central Florida
Orlando, Florida

This database analysis was conducted to determine which factors, or combination of factors, play a part in creating an unwanted outcome due to surprising or unexpected events encountered by pilots. The purpose of this study was to identify likely precursors to perceived surprising and unexpected events and, to advance our understanding of the overt behaviors and misbehaviors found in response to these events. This study also sought to determine if there were any significant differences between commercial air carrier and general aviation flight operations in regard to surprising and unexpected events. The results of this study indicated that the involvement of surprise or unexpectedness can indeed have a detrimental effect on the outcome of the flight. We also found indicators of the processes and mechanisms leading from surprise to an unwanted outcome.

Introduction

It is widely accepted that aircraft accidents, incidents, and events can result from novel and catastrophic unexpected situations. One need only look back at recent history for examples of fatal accidents that resulted from highly unusual situations such as US Airway's Flight 427 near Pittsburgh (National Transportation Safety Board, 1999), Alaska Airlines Flight 261 off the coast of Southern California (National Transportation Safety Board, 2000), and American Airlines Flight 587 in New York (National Transportation Safety Board, 2001). However, Kochan, Breiter, and Jentsch (2004) found surprising and unexpected events need not be rare, unusual, or catastrophic. Quite the opposite, pilots often describe normal, everyday occurrences as surprising or unexpected. They concluded that typical or normal events that occur daily (or nightly) in flight operations can also be surprising or unexpected to pilots.

This study builds upon previous research (Kochan, Breiter, & Jentsch, 2004) that identified what factors, conditions, and situations pilots and other users of the National Airspace System regard as surprising or unexpected. With this database analysis, we sought to deepen our understanding of surprise and unexpectedness by determining what underlying factors are present in situations that pilots deem surprising or unexpected. We asked: Are there certain factors, or a combination of factors, that are more likely to be present in situations where an unexpected event has a harmful effect on the outcome of the flight? Can seemingly trivial, everyday events, if surprising or unexpected to the pilot, produce an unwanted outcome?

Background

Research suggests that a person becomes surprised when their expectations are inconsistent with reality by an amount higher than could reasonably have been expected from the cues available and utilized by the individual (Kochan, Breiter, & Jentsch, 2004; Reisenzein, 1999). Expert pilots are normally able to process large amounts of information quickly and accurately, while continually and seamlessly modifying their situation awareness (Endsley, 2001; Wickens, 2002; Orasanu & Martin, 1998). However, decision making performance might be impaired, when pilots are confronted with events that do not adhere to expected schemata (Wickens, 2002; Endsley, 2001). Unexpected or surprising events may cause a disruption in cognitive processes (Reisenzein, 1999) leading to a decision making delay. This decision making delay lasts until the inconsistency, between what was expected and reality, is analyzed and integrated into the pilot's situation awareness (Meyer, Reisenzein, & Schützwohl, 1997). Reisenzein (1999) found that the more unexpected an event, the more significant this disruption in cognitive processes. The extent to which this potential interruption of cognitive processes occurs in the task of flying may influence the outcome of a particular maneuver or even the entire flight. It was with this assumption that the following database analyses were performed.

Purpose of Study

The purpose of this study was to determine (a) what factors or combination of factors are present in unexpected and surprising events; (b) find out to what extent these factors influence the surprising or unexpected event; (c) discover if there is a relationship between types of factors and severity of outcomes ranging from merely an event to a fatal accident; and (d) determine if there are any significant differences between commercial air carrier and general aviation flight operations, in regard to surprising and unexpected events.

Method

Search Procedure and Databases

Four databases (Table 1) were electronically keyword searched for the words "surprise" and "unexpected." Reports not relevant to this study were discarded. For example, if a reporter stated that, "it was not unexpected that…" or, if the reporter communicating the surprise or unexpectedness was not involved in the reported event (e.g. they were not a user of the National Airspace System) then the report was discarded. Also, reports submitted more than once were not included in the analysis. This study analyzed 638 reports.

Table 1. *Databases Reviewed for this Study.*

Database	Report Dates	*N*
National Transportation Safety Board (NTSB) Accident Database	1/1/1999 to 12/31/2003	131
National Aeronautics & Space Administration (NASA) Aviation Safety Reporting System (ASRS)	1/1/1999 to 1/1/2004	424
Federal Aviation Administration Accident and Incident Database (AIDS)	1/1/1999 to 12/31/2003	30
Major Air Carrier Aviation Safety Action Program (ASAP)	12/3/2002 to 10/19/2004	53

NTSB and AIDS reports are created as a result of an accident or incident investigation. ASRS and ASAP reports are compiled through voluntary reporting programs.

Report Analysis Procedure

Two aviation psychology researchers holding civilian flight instructor certificates reviewed the reports. Each report was examined for 71 variables. The variables selected for investigation were chosen to ascertain the location and environmental conditions surrounding the reported event, the demographics and experience level of the reporter, the type of aircraft, type of flight operation and purpose of flight, the factors surrounding the surprising or unexpected event, and the effect of the surprising or unexpected event on the outcome of the flight. The results from each report were coded and recorded into the Statistical Package for the Social Sciences (SPSS) v. 11.5.

Factors Surrounding "Surprise" and "Unexpectedness"

In addition to collecting data regarding the background and conditions of each surprising or unexpected event, each report was also reviewed for the presence of 35 factors (Table 2) believed to be involved with surprising and unexpected events.

Table 2. *Factors Associated with Unexpected or Surprising Events.*

Other's Surprise at Pilot's Actions	Go-Around
Surprise at Own Actions	Holding
Other Crewmember Actions	Delays
Maintenance Actions	Wind Takeoff
Loadmaster Actions	Wind Enroute
Passenger Actions	Wind Landing
Air Traffic Control	Wake Turbulence
Illusion or Disorientation	In-flight Turbulence
Aircraft State	Low Visibility
Automation	Icing
System Status	Bugs or Birds
Fuel State	Other Aircraft – Taxi
Landing Gear Position	Other Aircraft – Departure
Aircraft Position	Other Aircraft – Enroute
Aircraft Alerting Device	Other Aircraft – Landing
Airport Construction	NOTAMs
Runway Change	Temporary Flight Restrictions
	Fatigue

These contextual factors were selected for analysis because past research found them to be associated with surprising and unexpected events (Kochan, Breiter, & Jentsch, 2004). They were also selected because of their historical and reoccurring involvement in aviation events, incidents, and accidents. The task of this study was to determine if relationship exists between these factors, or a combination of these factors, and the manifestation of surprise and unexpectedness.

The Effect of Surprise and Unexpectedness

An important aspect of this study was to determine *to what extent* surprise and unexpectedness contributes to aviation events, incidents, and accidents. In this regard, each report was analyzed to discover what relationship existed between the involvement of surprise or unexpectedness and the outcome of the resulting event, incident, or accident. Following Helmreich, Klinect, & Wilhelm's (2001) model of threat and error management, if the report of surprise or unexpectedness had *no effect* on the outcome, then the surprise or unexpectedness was deemed *inconsequential*. If the report of surprise or unexpectedness *had an impact* on the outcome, then the surprise or unexpectedness was determined to be *consequential*. If the report of surprise or unexpectedness had a worsening effect on the outcome, then the surprise or unexpectedness was recorded as having *exacerbated* the situation.

The surprise or unexpected occurrences were then evaluated for their impact on the *outcome* of the flight; normal, reportable event (no damage or injuries), incident (damage and/or injuries less than accident threshold), or accident (substantial damage and/or significant injuries).

Results

A thorough look at these data indicated that the factors did not correlate adequately to perform a factor analysis. The factors and their frequency and percent occurrence in the reports are listed in Table 3.

Table 3. *Factors Most Frequently Involved with Surprising and Unexpected Events by Frequency and Percentage of Reviewed Reports (n=638).*

Factor	Frequency Present	Percent Present
Aircraft Position	420	65.8
Air Traffic Control	326	51.1
Other Crewmember Actions	270	42.3
Aircraft State	202	31.7
System Status	123	19.3
Automation	95	14.9
Inflight Turbulence	74	11.6
Low Visibility	64	10.0
Delays	62	9.7
Airport Contstruction	60	9.4
Other Aircraft - Enroute	60	9.4

Chi-Square tests for independence were conducted to evaluate the differences between the results of the unexpected or surprising event and the outcome of flight as displayed in Table 4.

Table 4. *Event Outcome vs. Flight Outcome by Percent within Outcome of Flight (n=638).*

Result of Event	Outcome of Flight			
	Normal	Event	Incident	Accident
Inconsequential	8.2	10.5	21.5	9.2
Consequential	21.3	34.0	18.5	18.4
Exacerbated	70.5	55.5	60.0	72.4

Chi-Square tests for independence were also conducted to find which factors involved in unexpected and surprising events were significantly different between events, incidents, and accidents, displayed in Table 5.

Table 5. *Relationship between Factors and Severity of Outcome Ranked by Strength of Association.*

Factor	Flight Outcome Severity Percent Present		
	Event	Incident	Accident
Air Traffic Control $X^2(3, 634=159.38)$, $p < .001$ (Φ=.501)	53.8	15.4	5.7
Wind Landing $X^2(3, 634=107.80)$, $p < .001$ (Φ =.412)	1.3	29.2	26.4
Other Crewmember Actions $X^2(3, 634=53.64)$, $p < .001$ (Φ =.291)	37.8	49.2	12.6
Automation $X^2(3, 634=36.01)$, $p < .001$ (Φ =.238)	15.1	3.1	0.0
Inflight Turbulence $X^2(3, 634=32.59)$, $p < .001$ (Φ =.227)	6.7	32.3	11.5
Aircraft Position $X^2(3, 633=19.48)$, $p < .001$ (Φ =.175)	62.9	56.9	54.0
Aircraft Alerting Device $X^2(3, 634=16.48)$, $p = .001$ (Φ =.161)	12.6	13.8	0.0
Other's Surprise at Pilot's Actions $X^2(3, 633=15.61)$, $p = .001$ (Φ =.157)	6.3	10.8	18.4
Maintenance Actions $X^2(3, 634=14.29)$, $p = .003$ (Φ =.150)	10.9	16.9	6.9
Other Aircraft – Enroute $X^2(3, 634=13.45)$, $p = .004$ (Φ =.146)	12.6	1.5	2.3
Illusion or Disorientation $X^2(3, 634=13.09)$, $p = .004$ (Φ =.144)	5.5	3.1	4.6

A Chi-Square two-way contingency table analysis was conducted to evaluate which factors involved in unexpected and surprising events were significantly different between air carrier and general aviation. The results of these analyses are displayed in Table 6.

Table 6. *Differences in Factors in General Aviation vs. Air Carrier Operations.*

Factor	Percent Present	
	Air Carrier	General Aviation
Wind Landing $X^2(1, 631{=}30.0)$, $p < .001$	2.4	14.2
Aircraft Position $X^2(1, 630{=}19.34)$, $p < .001$	57.9	74.6
Temporary Flight Restrictions $X^2(1, 631{=}13.11)$, $p < .001$	0.6	5.4
Wind Takeoff $X^2(1, 631{=}9.84)$, $p = .002$	0.3	3.7
Wind Enroute $X^2(1, 631{=}11.54)$, $p = .001$	1.2	6.1
Other's Surprise at Pilot's Actions $X^2(1, 630{=}7.04)$, $p = .006$	5.4	11.5
Illusion or Disorientation $X^2(1, 631{=}7.95)$, $p = .005$	5.1	11.1
Landing Gear Position $X^2(1, 631{=}5.64)$, $p = .018$	1.5	4.7
Airport Construction $X^2(1, 631{=}5.80)$, $p = .016$	6.9	12.5
Other Aircraft – Departure $X^2(1, 629{=}3.99)$, $p = .046$	3.3	6.8
Holding $X^2(1, 630{=}4.51)$, $p = .034$	3.0	0.7
Aircraft State $X^2(1, 631{=}5.18)$, $p = .023$	35.8	27.4
Aircraft Alerting Device $X^2(1, 631{=}8.04)$, $p = .005$	11.3	5.1
Inflight Turbulence $X^2(1, 631{=}11.56)$, $p = .001$	15.8	7.1
Automation $X^2(1, 631{=}16.44)$, $p < .001$	20.3	8.8
Other Crewmember Actions $X^2(1, 631{=}91.52)$, $p < .001$	60.0	22.3

Discussion

This study revealed that there is a relationship between the involvement of a surprising or unexpected event and the severity of the outcome of the flight (Table 4). In 72.4% of the accidents reviewed for this study, the involvement of surprise or unexpectedness did exacerbate the situation. On the other hand, the surprising or unexpected event was found to be inconsequential in only 9.2% of the accidents. We can see from Table 4 that in all 'outcome of flight' categories the surprising event was more likely to exacerbate the situation than not. Therefore, regardless of the ultimate outcome of the flight, surprise very often has a worsening effect on the situation. Interestingly, Table 4 shows that in 70.5% of the surprising or unexpected events that resulted in a *normal* outcome, the surprise or unexpectedness also exacerbated the situation. This category represents situations where surprise worsened the situation, but the flight continued normally never having crossed the event, incident, or accident threshold. This suggests that the occurrence of surprising or unexpected events might be a more nominal part of flight operations than previously thought. It is likely that the vast majority of surprising or unexpected events that end in normal outcomes go unreported.

This study found several factors which tend to be involved in more severe (incident or accident) flight outcomes (Table 5). It is interesting that the factor 'other crewmembers actions' is strongly associated with more severe flight outcomes. As would be expected, further analysis of this factor revealed it is more strongly associated with air carrier than general aviation operations. Further study into this area is needed to determine the nature, extent, and implications of the problem.

Results of this study also indicated that there are many types of surprising events in aviation. The fact that there are no consistent patterns of these events occurring suggests that potentially any event or combination of events can produce a situation which can end in an unwanted outcome as exhibited in the following examples.

Aircraft Position and Confounding Events. Findings from this study support research by Hoeft, Kochan, and Jentsch (2005) which revealed the flawed nature of the current NOTAM system. In this study, pilots repeatedly described the NOTAM system, which disseminates Temporary Flight Restriction (TFR) information, as unclear and difficult to use. This study found that TFRs are more of a general aviation problem than an air carrier problem (Table 6). In the example below, the pilot was conscience of nearby restricted airspace. However, an unexpected system malfunction contributed to the pilot's loss of awareness of the aircraft position and inadvertent penetration of a TFR. Aircraft Position was the most frequently (65.8%) occurring factor in all of the reports (Table 3).

They [ATC] were extremely busy and, I believe, were working another plane with a call sign of X, but I thought I heard a clearance. Near this time I experienced an ***unexpected*** *overload on my electrical system and had to flip the battery switch to correct it. This required me to reprogram my GPS which contains the communications I was using. I was unable to reach Orlando approach again and called Kissimmee tower. I had veered west and was attempting to circle south of the Disney World TFR and come back to the approach on runway 15, the runway in use; I was in contact with Kissimmee tower as I joined the approach. If I violated the TFR it must have been at this point. I was cleared to land* *by Kissimmee tower and landed on runway 15. I was advised that I had violated the Disney World TFR. (ASRS Report Number 578835 – Event)*

Air Traffic Control Actions and Landing Traffic. ATC instructions or actions were found to be a factor in 326 or 51.1% of the reports reviewed for this study (Table 3). ATC instructions were more likely to be involved in events than incidents or accidents (Table 5). The following is an excerpt from an ASRS report submitted by a corporate jet pilot describing a hurried departure in marginal weather at a busy airport. Note that the controller advised the crew to be ready for takeoff. A takeoff clearance while holding for departure usually would not be regarded as surprising. However, after analyzing their situation this crew was convinced that an immediate takeoff was unlikely and therefore were "surprised" and rushed into a potentially dangerous departure.

*The controller advised us to be ready to go. We acknowledged ok. And then, there was about a 3 minute break in the arriving traffic. Nothing happened. No takeoff clearance. We were spring loaded to go and then nothing happened. Finally, out of the clouds pops another aircraft on final. As I watched him get closer, I realized that we weren't going to be released. I relaxed, my copilot relaxed. Big mistake. Tower cleared us for an immediate takeoff. You can't even begin to imagine our total **surprise**. Both crew and engines weren't spooled up to go. As we were turning the corner for a rolling takeoff, tower comes back and asked if we were rolling! As soon as we replied affirmative, the controller sent the arriving aircraft around. The controller wasn't happy, the arrival wasn't happy and I wasn't happy. (ASRS Report Number 598909 – Event)*

Going-Around the Automation. The go-around is a maneuver intended to be used when an approach or landing needs to be discontinued. By its very nature, a go-around is not generally a pre-planned maneuver. However, pilots should be prepared for a go-around at any point during an approach (Federal Aviation Administration, 1999). Go-around was found to be a factor in 52 or 8.2% of the reports. Automation issues found in 14.9% of the cases (95) combined with the go-around maneuver increases the surprise factor. The following ASRS excerpt reveals that flight crews are not always ready to perform a go-around and an unexpected go-around can result in potentially dangerous situations.

*The aircraft was stabilized on approach and spacing with other traffic appeared to be comfortably spaced on TCAS II. Crossing the FAF at 2800 feet, the tower controller issued a clearance to climb to 4000 feet and to turn left to 360 degrees. I did not hear clearly the call sign on the clearance. I looked to the FO and asked him to verify the clearance being for us. My hands were on the flight controls as I was 'following' the autopilot on the approach. As the FO verified the clearance, I selected 'TOGA' mode of flight automation and proceeded with the normal GAR [go-around] callouts. Selecting TOGA automatically disconnected the autopilot and established nearly full power on both engines. As I was not looking directly at the flight instruments when selecting TOGA, the very rapid increase of power caused the aircraft pitch to increase past the desired attitude of 15 degrees to an attitude of 20 degrees, or possibly slightly higher. Although I instinctively placed forward pressure on the flight controls to counter the rapid change in pitch, the pressure was insufficient to stop the pitch at the desired attitude. In an attempt to smoothly lower the nose in the interest of passenger comfort, the aircraft experienced a 1 or 2 second stick shaker warning as we leveled at 4000 feet. Contributing factors: 1) An **unexpected** condition: an **unexpected** GAR at an **unexpected** phase of flight, 2) automation which contributes to large **surprise** factor: large and rapid power change in engines well below the wing creating an instant pitch change, and then disconnecting the autopilot. 3) The selection of TOGA at a time when concentration was not firmly established on flight instruments. (ASRS Report Number 575644 – Event)*

Sixteen factors were found to have significantly different rates of occurrence in general aviation and air carrier operations (see Table 6). Wind on landing was the most influential, significant factor between general aviation and air carrier operations (x^2 (1,631) = 30.0 *p*, <.001). This finding supports a recent study by the FAA which reported that wind accounted for 46.3 percent of all the FAR Part 91 weather related accidents between 1991 and 2001 (Federal Aviation Administration, 2001). In addition, wind on landing was strongly associated with more severe flight outcomes (Table 5).

Other Aircraft. Pilots often cited the sudden presence of other aircraft as surprising or unexpected. Other aircraft were a factor in 168 or 27.6% of the reports. Often poor traffic scanning on the part of the pilot contributes to situations where other aircraft appear suddenly. Interestingly, the presence of other aircraft enroute was a factor in 60 or 9.9% of the reports, more than any other phase of flight. In the ASRS report below the pilot was busy configuring his aircraft for departure and not focusing his attention outside the aircraft while taxiing for takeoff.

While completing the [before takeoff] checklist, I heard the aircraft calling ground stating that they were clearing runway 19 onto taxiway D, which was ideal for him since taxiway D leads into the airline terminal ramp. As I looked up in ***surprise*** *and shock (at that moment I realized I didn't check the runway for traffic), the ERJ was turning off with all exterior lights still on, which caused temporary blindness. (ASRS Report Number 598235 – Event)*

Conclusions

This study established that surprising and unexpected events can and do have a negative effect on the outcome of flight. We found several factors that are consistently involved in surprising and unexpected events. We also determined what flight outcome each factor is likely to be involved with. However, a simple formula explaining what combination of factors, are more likely to cause a surprising or unexpected event resulting in an unwanted outcome remains allusive. Perhaps the most important finding from this study is that potentially any factor or combination of factors can create a surprising or unexpected event that leads to an unwanted outcome.

Acknowledgements

The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of Transportation, Federal Aviation Administration, or the University of Central Florida. This study was sponsored, in part, by FAA Grant 99-G-047, Dr. Eleana Edens, grant monitor.

References

Endsley, M.R. (2001). Designing for Situation Awareness in Complex System. *In Proceedings of the Second International Workshop on symbiosis of humans, artifacts and environment, Kyoto, Japan.*

Federal Aviation Administration (2001). *NASDAC Review of NTSB Weather-Related Accidents.* Retrieved January 24, 2005 from https://www.nasdac.faa.gov/aviation_studies/weather_study/studyindex.html

Federal Aviation Administration. (1999). *Airplane Flying Handbook.* U.S. Department of Transportation.

Helmreich, R. L., Klinect, J. R. & Wilhelm, J. A. (2001). Models of threat, error, and CRM in flight operations. In *Proceedings of the Tenth International Symposium on Aviation Psychology*, Columbus, OH: The Ohio State University

Hoeft, R. M., Kochan, J. A., & Jentsch, F. (2005). A human factors analysis of the current U.S. notices to airman (NOTAM) System. *The International Journal of Aviation Psychology.* 15(1), 91-109.

Kochan, J. A., Breiter, E. G., & Jentsch, F. (2004). Surprise and unexpectedness in flying: Database reviews and analyses. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting.* New Orleans, LA: Human Factors and Ergonomics Society.

Meyer, W., Reisenzein, R., & Schützwohl, A. (1997). Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion*, 21(3), 251-274.

National Aviation Safety Data Analysis Center (n.d.a). *Aviation safety reporting system.* Retrieved October 11, 2003 from http://nasdac.gov

National Aviation Safety Data Analysis Center (n.d.b.). *Accident and incident databases.* Retrieved October 11, 2003 from http://nasdac.gov

National Transportation Safety Board (1999). *USAir Flight 427 Aircraft Accident Report.* Retrieved February 3, 2005 from http://www.ntsb.gov/publictn/A_Acc2.htm

National Transportation Safety Board (2000). *Alaska Airlines Flight 261 Aircraft Accident Report.* Retrieved February 3, 2005 from http://www.ntsb.gov/events/2000/aka261/default.htm

National Transportation Safety Board (2001). *American Airlines Flight 587 Aircraft Accident Report.* Retrieved February 3, 2005 from http://www.ntsb.gov/events/2001/AA587/default.htm

National Transportation Safety Board (n.d.). *Aviation accident database and synopses.* Retrieved October 11, 2003 from http://www.ntsb.gov/

Orasanu, J., & Martin, L. (1998). Errors in Aviation Decision Making: A factor in Accidents and Incidents.Retrieved December 26, 2004 from www.dcs.gla.ac.uk/~johnson/papers/seattle_hessd/judithlynne-p.pdf

Reisenzein, R. (1999). The Subjective Experience of Surprise. Chapter prepared for: H. Bless and J.P. Forgas (Eds.), *Subjective experience in social cognition and social behavior*. Philadelphia, PA: Psychology Press.

Wickens, C.D. (2002). Situation Awareness and Workload in Aviation. *Current Directions in Psychological Science.* Blackwell Publishing Inc.

PROGRAM UPDATE AND PROSPECTS FOR IN-FLIGHT SIMULATION UPSET RECOVERY TRAINING

Janeen A. Kochan
James E. Priest
General Dynamics
Roswell, NM

The Flight Research Training Center, established in 2002 in cooperation with the Federal Aviation Administration, focuses on improving the safety of commercial air transportation through the reduction of the loss-of-control events, which continue to be the leading cause of fatal commercial air carrier accidents (Boeing Commercial Airplanes Group, 2004). The primary research purpose of this program is the optimization of in-flight simulation based upset recovery training. The goal of the training is to have a beneficial impact on the loss-of-control accident and incident rate. The program is designed to collect research data through an extensive training program offered to commercial airline pilots. To date, more than 235 commercial pilots have completed the integrated two-day program which includes classroom, aerobatic aircraft, and advanced in-flight simulation aircraft training on how to best respond to a variety of upset situations. This paper presents the results of the data collection and analysis effort for the FAA-Upset Recovery Training (URT) program for the twenty-four month period from August 8, 2002 through July 30, 2004.

Introduction

Program Background

The fundamental goals of the FAA-URT project are (a) to conduct research to optimize in-flight simulation (IFS) based upset recovery training, (b) meet the pilot training needs of commercial air carriers, (c) to design and develop IFS technology and systems specifically for the URT role, and (d) to have a beneficial impact on the loss-of-control accident, incident, and event rate. This results in a program which is hybrid in nature creating two complementary, yet independent, activities: operational training and an empirical study. Quasi-experimental field studies inherently present a host of research obstacles (Shadish, Cook, & Campbell, 2002) which are amplified in this blended training and research arena. A balance had to be reached between the need for efficient training operations and the need to collect sufficient data to provide a basis for the optimization activity.

Original Training Protocol

The structure and deployment of the training protocol has been continually monitored and revised during the course of the study based on participant feedback. For the first eighteen months of the study (August 8, 2002 through December 31, 2003) the training protocol was composed of three modules and conducted over a two day period. The first module was a classroom lecture where participants received instruction in causes of upsets, aerodynamic fundamentals, and recovery techniques. The second module was usually a training flight in the Aerobatic Bonanza. The Bonanza flight exposed the pilot to general aircraft characteristics, G-force awareness, slow flight and stall awareness, limited aerobatics, and unusual attitude recoveries. The third module was the IFS Learjet aircraft training where the participant experienced real-world upset events and practiced various recovery techniques. This module began with a flight rehearsal session using a ground-based (non-motion) simulator

Protocol Modifications

Changes in the program protocol based on feedback from the first 201 participants of the program were made beginning in January, 2004. Evaluation forms from the participants and instructor comments regarding the structure of the program and the usefulness of each of the training elements were reviewed. Based on the high frequency of comments regarding the order of events in the training, the structure of the program was changed. This involved separating the classroom briefings into two sessions; one prior to the Bonanza flight and one prior to the ground simulator and Learjet flight. Strict adherence to flying the Bonanza prior to the Learjet was also implemented in response to comments from the participants. These modifications to the protocol changed the order of the presentation of elements; however, the content of each module was not substantially altered.

In-Flight Simulator Learjet Flight

The URT protocol is an integrated, multi-part training event. However, the majority of the measures during the first 24-months of the program have focused on the efficacy of the In-Flight Simulator

Learjet training flight. A typical flight consisted of five phases: (a) familiarization exercises, (b) beginning evaluation exercises, (c) "g" awareness and confidence maneuvers, (d) upset recovery practice events, (e) ending evaluation exercises. The Learjet IFS aircraft, pre-programmed with upset events, is used to teach actual upset recoveries. The events programmed into the simulation system range from atmospheric effects and a wake turbulence encounter to extreme control failures and control surface hardovers. The simulation was of a light-to-medium size transport aircraft that is near max gross weight so that the inertias produce near worst case handling qualities.

Quantifying the Training Effectiveness

To optimize the in-flight simulation based upset recovery training, we needed to be able to measure how much the participant's ability to recover, from a variety of upsets, improved during the training. We also needed to assess the value of the various events to the participant. Our initial research questions were (a) how much did the participant learn from the URT experience, and (b) what elements did the participant find most useful and why?

Recovery Ratings

Measuring a pilot's ability to recover is a difficult task. Unfortunately, the seemingly straightforward concept of measuring performance parameters such as reaction times, maximum bank or pitch angles, etc. do not provide an accurate measure of a pilot's ability to recovery from any given unexpected event. The essence of the difficulty in assessing human performance in this task is captured in the following example.

Consider that a single driver has two cars; car "A" steers poorly, car "B" steers like a dream. If you follow each of these cars for 10 miles, with the same driver, you may not be able to tell which car drives the best. When car "A" is driven, the driver pays strict attention to the task and rarely strays from the center of the lane. When car "B" is driven, the driver's attention may wander to other things resulting in straying further from the center of the lane than occurred with car "A". To the outside observer, using quantitative measures, it might appear that car "A" handles better. However, the most expeditious (and perhaps accurate) way of finding out which car drove the best is to ask the driver who will be able to tell you unequivocally about the (a) mental and physical workload, (b) level of apprehension and or stress, and (c) confidence experienced in performing the task. In this example, driver opinion would say car "B" performed better. Thus, in the long run, it may be much more cost effective and accurate to ask the driver to provide the performance evaluation.

Measuring the quality of a pilot's recoveries to upset events presents a problem similar to the driving task. We must consider both the perceptual-motor performance, physical and mental workload, and the level of confidence one has in responding to an upset. Flight test organizations around the world have adopted the Cooper-Harper rating scale to facilitate quantifying aircraft handling qualities (Gawron, 2000). The Cooper-Harper scale incorporates performance and workload measures to assist an evaluation pilot in determining a single rating of the handling qualities of a particular aircraft.

The original Cooper-Harper Handling Quality Scale has been adapted to fit the needs of the URT program. The Recovery Rating Scale (RRS) is administered near the beginning of the flight to obtain a "beginning" rating and then again after practice near the end of the flight for an "ending" rating. These "beginning" and "ending" ratings did not intend to measure what had been learned from the entire course. Instead, the purpose of the scale was to help determine how much the participant learned specifically from the in-flight simulation upset recovery practice.

During this initial phase of the program, no effort was made to measure the amount learned in the entire course, or how much each element contributed to the overall program. However, participants did have an opportunity to comment on their perceptions of the elements and the benefit of the overall course as part of the post flight evaluation form.

Participant's Evaluations of the Program

The second question of interest, how valuable was the course to the participant, was addressed by a postflight evaluation form. This form contained specific, liker-scaled and open-ended questions regarding the participant's perception of each element of the training protocol.

Research Questions

Our initial research questions were (a) how much did the participant learn from the URT experience, and (b) what elements did the participant find most useful and why? To address these questions, we posed the following specific questions to guide our initial analysis of the data:

- Is there any significant improvement in the Ending RRS scores over the Beginning RRS scores?
- What is the relationship between total flight time and the Beginning RRS scores?
- What is the relationship between total flight time and the Ending RRS scores?
- What other factors influence the Beginning RRS scores, Ending RRS scores, or the magnitude of the difference between them?
    - What effect does military training have on the RRS scores?
    - What effect does previous aerobatic experience have on the scores?
    - What effect does being an instructor pilot have on the scores?
- What are the participants' perceptions of the URT program?

Method

Participants

The participants to date were 248 volunteers recruited by direct contact to airline training departments, website solicitation, and word of mouth. Participants were also informed of the study through numerous articles written about the project and published in aviation journals and trade magazines. Program contact information for participants was often included in the reports and articles.

Data analysis for this report was completed using data sets from 185 qualified air carrier pilots representing 27 different U.S. Part 121 air carriers. The additional participants (not included in these analyses) were from government organizations (e.g., FAA and NTSB), universities, research facilities, and private organizations (e.g. Airline Pilots Association, National Business Aircraft Association, etc). The exclusion of these data facilitated a focused look at the representative air carrier pilot.

Study participants included three females and 182 males. Approximately one-third (68) of the participants had military training and 121 participants had experience as either military or civilian instructor pilots. All participants held at least an FAA Commercial pilot certificate with an Instrument Rating, although the majority (157) held an Airline Transport Pilot certificate. All participants maintained a current FAA Medical Certificate.

Data Collection

Data was collected by the program administrator and Safety Pilots through forms, questionnaires, and instructor notes. All materials containing study materials and data were kept in secure quarters, accessible only to the study principals and researchers.

Results

Data Screening

All data in this study were next reviewed for accuracy of input into the SPSS file by checking for (a) out-of-range values, (b) plausible means and standard deviations, (c) univariate outliers, and (d) missing data. Pairwise plots for nonlinearity and heteroscedasticity were also reviewed when necessary for the statistical method used. When data was found to be missing, it was random in nature and therefore posed little threat to the validity of the results. Unless otherwise specified, all analyses were conducted with alpha level set at $p < .05$. Analyses were conducted using the Statistical Package for the Social Sciences v. 11.5. All graphic scales depicting data analysis are scaled identically for ease of comparison.

Inter-rater Reliability

The study was conducted and data were collected by six different instructor pilots (safety pilots). Analyses were conducted and no significant differences were found between Beginning or Ending RRS scores from participants of different instructor pilots $F(5, 144) = 2.175$, $p = .060$.

Upset Recovery Rating Score Differences

A paired-samples t-test was conducted to evaluate the impact of the in-flight simulator Learjet training on responses on the RRS scores. There was a statistically significant decrease (improvement in perceived performance) from the beginning rating scores ($m = 6.29$, $SD = 2.06$) to the ending rating ($m = 2.87$, $SD = 1.13$), $t(149) = 24.13$, $p < .0005$. The eta squared statistic (.80) indicated a very large effect size.

Effects of Total Flight Hours on Upset Recovery Rating Scores

A repeated-measures univariate analysis of variance (ANOVA) was conducted to investigate the effects of total flight hours on Beginning and Ending RRS scores. The number of participants per cell was not equal, *n* ranging from 25 to 35 per cell as depicted in Table 1.

Table 1. *Mean and Standard Deviations of RRS scores Participants by Total Flight Hours*

Flight Time	n	Beginning RRS (*SD*)	Ending RRS (*SD*)
< 5,000	35	7.03 (2.24)	2.69 (1.20)
5,001 – 10,000	42	6.24 (2.07)	2.69 (1.20)
10,001 – 15,000	45	5.98 (1.92)	2.82 (1.03)
> 15,000	25	5.76 (1.76)	2.56 (0.87)

A significant difference was found for the main effect of flight time on the RRS scores, *F* (3,143) = 3.11, *p* < .05 with a moderate effect size (partial Eta squared = .06). Estimated marginal means and standard errors were evaluated post-hoc (*a posteriori*) for differences using the Least Significant Differences pairwise multiple comparison test. No significant interaction effects with the dependent variable (RRS scores) were found. There were significant differences (*p* < .05) in the means of the beginning scores of the lowest time pilots as compared to each of the other three groups (5,001 – 10,000; 10,001 – 15,000; > 15,000). A graphical depiction of the effects of total flight time experience is displayed in Figure 1.

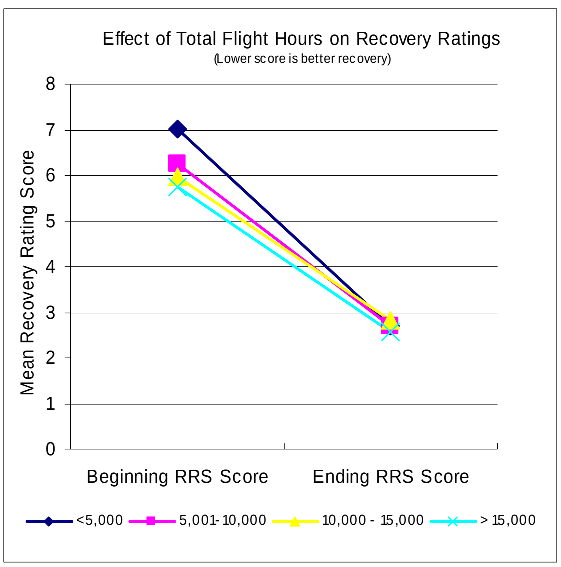


Figure 1. Effect of total flight hours on upset recovery improvement.

Effects of Type of Training on Upset Recovery Rating Scores

A repeated-measures univariate analysis of variance (ANOVA) was conducted to investigate the effects of military training on Beginning and Ending RRS scores. The number of participants per cell was not equal as displayed in Table 2. A significant difference was found for the effect of type of training on the RRS scores, *F* (1,145) = 4.41, *p* < .05 with a small effect size (partial Eta squared = .04) as presented in Table 2.

Table 2. *Mean and Standard Deviations for Different Types of Training*

Type of Training	*n*	Beginning RRS (*SD*)	Ending RRS (*SD*)
Civilian Only	93	6.46 (2.21)	3.01 (1.23)
Military and Civilian	54	5.93 (1.71)	2.55 (0.79)

Effects of Aerobatic Experience on Upset Recovery Rating Scores

A repeated-measures univariate analysis of variance (ANOVA) was conducted to investigate the effects of previous aerobatic training on Beginning and Ending RRS scores. No aerobatic training was indicated by "None" while recreational or minimal aerobatic training was designated "Some". "Extensive" aerobatic training was either (former) military pilots or those who had performed in airshows.

A significant Levene's statistic (*p* < .05) was found for the Beginning and Ending RRS scores; therefore ,corrections to the alpha level were made for these analyses. The means and standard deviations for each aerobatic experience group are shown in Table 4.

A significant difference was found for the main effect of previous aerobatic training on the RRS scores, *F* (2,146) = 4.71, *p* = .01 with a moderate effect size (partial Eta squared = .06). Estimated marginal means and standard errors were evaluated post-hoc for differences using the Least Significant Differences pairwise multiple comparison test. No significant interaction effects with the dependent variable (RRS scores) were found. There were significant differences (*p* < .005) in the estimated marginal means of the pilots with no aerobatic experience, and those with extensive aerobatic experience for beginning RRS scores.

Table 3. *Mean and Standard Deviations of RRS scores for Participants by Aerobatic Experience*

Aerobatic Experience	*n*	Beginning RRS (*SD*)	Ending RRS (*SD*)
None	36	6.89 (2.35)	3.33 (1.49)
Some	51	6.29 (2.14)	2.88 (0.99)
Extensive	62	5.90 (1.71)	2.54 (0.84)

Effects of Flight Instructing on Upset Recovery Rating Scores

A repeated-measures univariate analysis of variance (ANOVA) was conducted to investigate the effects of flight instructing on Beginning and Ending RRS scores. No significant difference was found for the effect of flight instructing on the RRSs, F (1,148) = .317, p = .57 as shown in Table 4.

Table 6. *Mean and Standard Deviations for Flight Instructors and Non-Flight Instructors*

Flight Instructor	*n*	Beginning RRS (*SD*)	Ending RRS (*SD*)
No	49	6.41 (2.20)	2.93 (1.18)
Yes	101	6.24 (1.99)	2.83 (1.10)

Participants' Perceptions of the Upset Recovery Training Program

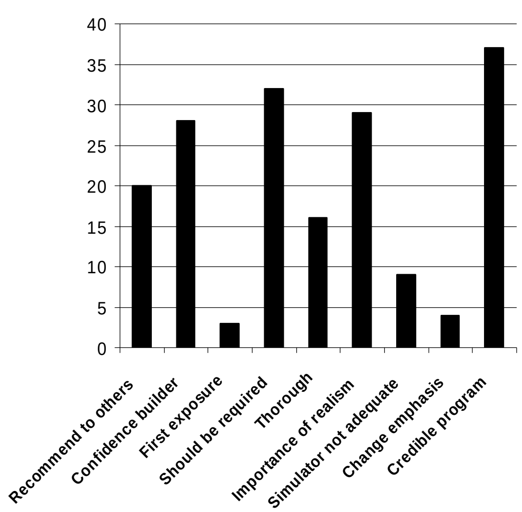


Figure 3. *Frequency and types of responses to open-ended question, "comments on course," from Upset Recovery Training course evaluations.*

Participants' perceptions of the URT program were evaluated through frequency analysis. A summary of these results are presented pictorially in Figure 2 which shows the frequency of overall comments from the participants in the study. In addition, the relative importance of each element of the in-flight simulator Learjet flight was determined by rank order (1 = fair to 5 = excellent) of the mean scores from the participants' course evaluation forms.

Discussion

Overall, these results suggest a strong positive influence of the Upset Recovery Training Program on a pilot's ability to respond to an inflight upset. Specifically, the RRS scores indicate a very strong training effect. It is interesting to note the effect of flight times on the Beginning and Ending RRS scores. Even though there is a significant main effect, the bulk of the variance between groups is with the lowest time pilots (< 5000 hours) and those pilots with > 5000 hours. This suggests that pilots of all experience levels (based on total flight time) gain essentially the same benefit from the training.

The effect of military training on the RRS scores is also worth noting. The effect size is particularly small and the majority of the variance was found to be in the beginning scores. Since many civilian pilots have not had the opportunity to perform aerobatics, it is not surprising that there is a significant effect of having experienced aerobatic flight on the RRS scores. Additional analyses were undertaken to determine how much of this effect was confounded by military training. When the effects of aerobatic experience were held constant, there were virtually no differences in the two groups.

The lack of a significant effect of experience as a flight instructor on RRS scores is not particularly surprising. This specialized, advanced airmanship type regimen is not currently taught to flight instructors, therefore it is also not taught *by* flight instructors (Federal Aviation Administration, 2002). Furthermore, most airline pilots no longer participate in instruction outside of the airline.

Although this adaptation of the Cooper-Harper Scale (RRS) has not been previously validated as a measurement instrument, these results offer some interesting insights into its usefulness. First, the six instructor pilots assisting in the use of the scale found homogeneity in their participants' Beginning and Ending RRS scores. This suggests that the scale is being used in a consistent manner across all users. Next, the RRS scores follow known trends in pilot expertise research for total flight time (Jensen, 1995), whereas the lower time pilots showed significantly higher Beginning RRS scores than the higher time pilots. Finally, the RRS scores followed the hypothesis that pilots with aerobatic training would have lower RRS scores (better performance) than those without any aerobatic training. Neither the total flight time, nor the aerobatic experience level of the participants was known to the instructor pilots before the training event minimizing experimenter bias. Therefore, one could conclude that the RRS is a valid measure of this task.

The participants' perceptions of the program appear very positive in these data; however, their scores reflect a ceiling effect on their ratings of the specific course elements.

Limitations of Current Study

There were numerous limitations in the current research avenue of this hybrid training-research program. The most salient issues (which we are already addressing) were:

- The study population was self-selected volunteers who, in many cases, held positions in the airline's flight training department or management. Even though there were no significant effects of being a flight instructor, per se, on RRS scores, there is still a need for a more representative population to be able to better generalize the results of the study.

- Extraneous variance in the experimental setting and measurement techniques resulting mostly from the nature of aviation (weather, mechanical malfunctions, etc.) were present.

- The testing effect of a repeated measures design without a control or pretest condition.

- Mono-operation bias whereas the focus was on only one aspect of a multi-part training program. The influence of the academics, Bonanza flight, and ground simulator were not controlled for or measured in the current format of the study.

- The use of only one performance measurement technique created a mono-method bias which only measures one aspect of the construct of training effectiveness.

- The possible unreliability of the RRS where measurement error may have occurred as it has not been cross-validated or scrutinized as an accurate representation of pilot performance.

Conclusions and Future Direction

Future directions of the study will focus on addressing the identified limitations discussed above. New protocols, forms, and data collection efforts have been established and implemented as countermeasures to the threats to validity which can be controlled or measured. For example, instructor calibration and collection of more detailed flight time and pilot experience will aid in controlling potential confounding effects in the study. More exacting pilot demographics and flight times allow for control of experience levels. Measurement techniques have been established to specifically test for the training effects at each stage of the training. These enhancements to the research protocol will provide richer data from which the training program will ultimately benefit.

Acknowledgements

This research was supported, in part, by the U.S. Department of Transportation, Federal Aviation Administration Human Factors Research and Engineering Division, Contract Number DTFA01-02-C-00088, and the New Mexico Department of Transportation-Research Bureau. The authors wish to thank all of the pilots who have participated in the Flight Safety Research Center Upset Recovery Training Program.

References

Gawron, V. J. (2000). *Human performance measures handbook.* Mahwah, NJ: Lawrence Erlbaum Associates.

Federal Aviation Administration (2002). *Flight Instructor Practical Test Standards for Airplane with Change 1.* FAA-S-8081-6B. Washington, DC: Flight Standards Service.

Jensen, R. S. (1995). *Pilot judgment and crew resource management.* Aldershot: Ashgate Publishing Limited.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin Company.

REDEFINING CRM

J. Norman Komich
The Aviation Connection
55 Middlebury Lane
Beverly, MA 01915
978-927-9566
Email: jnkomich@aol.com

The recent revision (120-51E, dated 1/22/04) to the Crew Resource Management Training Advisory Circular failed to provide a specific definition of CRM. This void is an issue with those who agree with Montaigne when he observed "No wind favors the sailing ship without a destined port." Since its inception over twenty five years ago, CRM has undergone considerable evolution and the industry now finds itself in the seventh or eighth "Generation" of CRM training. Interestingly, in the very first Advisory Circular (120-51A) the following statement was made: "The essence of CRM training is to reduce error in the cockpit." In spite of that specific focus, during the last quarter of a century, CRM training has been whatever the program developer wanted it to be and the result has included such diverse subjects as Post Traumatic Stress, Security, Unruly Passengers, Scheduling Issues, and Uniform Codes. The original definition of CRM as "The effective utilization of all available resources including liveware, hardware and software, to achieve safe and efficient flight operations." was a worthy "goal" which unfortunately was more theoretical than practical; and no doubt contributed to why the current AC has no specific definition. Safety and efficiency do not always go hand in hand and therein lies the rub. *It is time the industry put the practical side of the issue first and then back that up with theory*. With that in mind, I make the observation that the industry has failed abysmally to take advantage of the huge resource of line pilot experience. Line pilots who achieves tens of thousands of hours accident and incident free has developed their own "bag of tricks" to stay out of trouble. Academicians, management pilots, and even union members, do NOT adequately represent the line pilot. With that in mind, I offer the following NEW and specific definition of CRM: "Cockpit Resource Management is the comprehensive utilization of all available resources including people, equipment and procedures, to attempt to get the job done correctly while staying out of trouble." There are an infinite number of ways to do this and each annual recurrent training should address some of those techniques. GAIN, ASAP, Line Pilot Reports, FAA violations, Accident and Incident Reports, and the ASRS reporting system are all excellent starting point to gather these techniques. Too much of that data is simply NOT making it to the cockpit. The industry must come to grips with the fact that with each new technological improvement, each new aircraft design and each now operational improvement, more challenges are being faced by the line pilot and CRM training is one way to aid the line pilot in coping with these challenges. Consequently, CRM training remains a journey and NOT a destination.

REDEFINING CRM

The twenty fifth anniversary since the first international workshop on CRM in 1979, recently passed without much fanfare. What was the reason for that lack of attention? Some might say it is a result of the fact that the aviation industry has adequately achieved what it set out to accomplish twenty-five years ago. Such an attitude is reinforced by the lack of specific CRM training that is taking place today. On the other hand some might believe that the lack of attention given to the passing of a quarter century is because the industry has failed to achieve ALL that it might have accomplished. When one considers the vast amount of time, effort and money that has gone into CRM training during this period, it is not difficult to understand why the industry is not celebrating such a lack of success. These are two very contrasting and contradictory points of view and they lay the foundation for this paper.

Redefining CRM must begin with addressing the need for a new definition. Why redefine CRM if the old definition is adequate?? Put another way, "Does the old definition suffice?" To answer those questions one must ask "What IS the old definition of CRM?" Anyone with any experience in the field is acutely familiar with the Mantra: "CRM is the effective utilization of all available resources including liveware, hardware, and software to achieve safe and efficient flight operations." That is certainly a worthy goal; simultaneously achieving safety and efficiency is the ultimate goal of ALL flight operations regardless of the mission. Airlines, military, corporate, air ambulances, off shore re-supply, etc., have very different missions but they all want to succeed in that mission and they all want to do so safely. No wonder the old definition has lasted as long as it has. The theory is rock solid. So how is the practical application?

To answer that question, one must ask how that practical application will be evaluated? One approach is to simply poll attendees of CRM training and ask if they feel safer than before the training. Another is to compare accident statistics from before and after CRM training. This author finds neither of those approaches acceptable for the following reasons. Self assessment is simply not objective enough and statistics remain too similar to the bikini: what they reveal is enticing but what they cover up is vital. A low accident rate tells NOTHING of what is routinely going on in the cockpit. Aside from the last thirty minutes of the cockpit voice recorder in an accident review, there are only two opportunities for the industry to be exposed to how pilot's function in the cockpit during day to day operations.

One is observing flight crews in the simulator. This author hates to rain on the parade of those who promote the simulator as the panacea for most of the aviation industry's safety ills, but sadly, they are kidding themselves and the industry. While the simulator is an excellent tool for teaching procedures, regardless of the amount of money spent on improving the high technology of motion and visual and regardless of the attempts to make LOFT (Line Oriented flight Training) provide the atmosphere of line operations, the fact remains that the pilot walks into the simulator and he knows that. In the simulator, one can NOT run out of gas, or be four hours late, or be rushing to beat a curfew, or have VIP's on board, or be actually fatigued after 14 hours on duty, or be hungry or whatever. Other characteristics which decrease the effectiveness of simulators include "glitches" in the software that have the simulator NOT responding like the real aircraft (negative G's being one of the more obvious) and the following war story in which the pilot was right in the middle of his LOFT scenario in which he had a key decision to make when the phone rang and the instructor began dialogue with another instructor as to where they were going out to eat that evening! Acknowledging these facts will provide a key cornerstone to understanding the efficacy (or lack thereof) of the practical application of CRM.

The other opportunity to experience the how practically CRM is being applied is through LOSA (Line Oriented Safety Audit) in which fellow crew members or researchers, ride along "just to observe" and note any errors made by the crew. It is pre-briefed that this audit will be "non-judgmental" and the crew is to act as if "they weren't there." Anyone who believes that the presence of an observer has NO impact on how the flight crew functions is VERY naïve. The Heisenberg theory empirically proved that the presence of one body has a definite impact on the movement of another body and nowhere is this truer than in the cockpit. With that acknowledgment, the other measurement of effectiveness of CRM is similarly diluted and that leaves the following and perhaps most legitimate gauge: line pilot observations.

NASA ASRS and ASAP reports provide only the tip of the iceberg. Regardless of the wording in the ASR and ASAP agreement, pilots are understandably not going to divulge all the nuances of the event in these reports. And how much is really learned from cursory facts?? The next best thing to "being there" is hearing about it "from the horses mouth. These stories are never made public for obvious reasons but any failure of the aviation industry to acknowledge them does and will leave a large void in the safety structure. And it is that void that precipitates the need for redefining CRM.

The industry has been all too quick to celebrate its successes and all too unwilling to acknowledge any failure. It is extremely ironic to note that some of those who are so quick to criticize a captain for not accepting input from his other crew members, are themselves too unwilling to accept any constructive criticism from others. This author is not suggesting that the industry has necessarily "failed" in its attempt to teach CRM but he does suggest that much more success might have been achieved; particularly when one considers the vast resources applied to CRM over the last twenty five years.

Having established the fact that while the theory of the old definition was sound, its practical application has been lacking, let us address another important reason for redefining CRM. The old definition of CRM made reference to "use of ALL available resources including liveware, hardware and software" and yet for the first sixteen years, all that was focused on was the "liveware" part of the equation. Retired American Airlines Captain, Bob Besco, pointed out that all that was being addressed was "Small Group Dynamics". There was no emphasis on aviation specific issues, just "how the crew got along." This myopic focus lead to the evolution of CRM from "Cockpit Resource Management" to "Crew Resource Management." When the researchers ran out of interactive issues between pilots, instead of addressing pilot-specific issues such as situational awareness, or CFIT (Controlled Flight Into Terrain) or Fuel Management, etc., they simply expanded the size of the "team" and focused on joint training with flight attendants, dispatchers and maintenance personnel.

The philosopher Montaigne once observed that “No wind favors the sailing ship without a destined port.” When the industry ran out of interactive “team” concepts, CRM training was indeed adrift at sea and subject matter for CRM training took on any appearance that the program developer desired. Post Traumatic Stress Syndrome, Security, Unruly Passengers, Scheduling Issues and Uniform compliance took the place of legitimate aviation safety issues. One glaring example of this lack of focus was discussion of a flight crew that encountered a severe wind shear that almost caused the airplane to crash. The entire focus of the class was on the after affects (PTSD) of the event on the crew. After listening to the description of the affect of the wind shear on the airframe, one pilot asked, “What angle of bank did the aircraft achieve before you were able to recover?” Not only was his legitimate question NOT answered, the CRM facilitator actually had the gall to declare “We will NOT discuss ANY airplane specific issues here! We are just discussing the success of the Critical Incident Response!” The absurdity of such a statement defies description. If only the industry would apply a fraction of focus on avoiding the accident that it does in intervening with the mishap crew, it would be in far better shape. Without a clear focus on an accurate definition of CRM, the industry will never achieve all the success in avoiding accidents that it might.

PhD’s were brought on board at many carriers to train these interactive skills but in some way, the industry was actually doing a disservice to its pilots by over emphasizing that as long as they “got along” and “communicated”, they were safe. While good interactive skills were “necessary” for a safe flight, they were by no means “sufficient”. The American Airlines accident at Cali was a glaring example that CRM training required much more than small group dynamics. The interactive skills of that particular crew were fine; they just suffered from a classic loss of situational awareness under high workload; one of the many safety concepts ignored while focusing on interactive training with other working groups.

The Cali crash resulted in significant backlash towards CRM training. After much prodding, the industry finally acknowledged that more than just small group dynamics was needed. At a pilot meeting at one of the major carriers, when the Chief Pilot made the statement that “CRM is dead at XYZ” the pilots cheered and gave him a standing ovation. At another major carrier, the “Next Generation” of CRM training was introduced with the statement “We are going to drain the hot tub in CRM!” While this was certainly a step in the right direction, acknowledging what NOT to do was still not enough. In the very first FAA Advisory Circular on CRM the following statement was made. “The essence of CRM training is to reduce error in the cockpit.” How should the industry accomplish that goal? After sixteen years of focusing almost exclusively on interactive skills to reduce error, Cali glaringly pointed out that the emperor has no clothes; and so sixteen years later, the industry finally began to focus on what it had failed to do so from the beginning.

While the intent was good, the result was abysmal. The industry again refused to get into specific aviation issues and instead came up with shallow concepts such as the Volant Model which basically advocates that “If you do everything right, you will not do anything wrong”. Malcom Armstrong, Director of Safety for one of the major carriers, succinctly shot holes in that model when he said “Most people do not come to work intending to have an accident. They are trying to do the right thing. Flawed training, improper priorities, and defective procedures are what lead to an accident.” The current Threat and Error Management Model is yet another generic attempt that fails to address specific aviation safety issues. Likewise, the goal of “Avoid, manage or mitigate the consequences of error” is yet another theory that sounds great until one attempts a practical application. I have asked many of its proponents for a specific example of “mitigating the consequences of an error” and I have yet to hear one legitimate one. Why didn’t the aviation community as a whole ask for such specific examples instead of blindly jumping on the bandwagon?

All of this is water under the bridge. None of the time, money, nor effort can ever be recaptured nor can any of the accidents that have occurred during that last twenty-five years be corrected. It is time for the industry to change by design rather than just by knee jerk reaction to yet more accidents and loss of life. The definition being proposed here begins with going back to the original concept of “Cockpit Resource Management” since the cockpit and aviation should be the focus of the training.

The new definition is given as “Cockpit Resource Management is the comprehensive utilization of all available resources including people, equipment and procedures, to attempt to get the job done correctly while staying out of trouble.” The emphasis on “attempt” is the practical acknowledgement that safety and efficiency do not always go hand in hand

and sometimes, the job will not always get done. Tenerife and Dryden have been held up as examples of poor CRM in literally thousands of CRM classes. What was the lesson learned from them? From the "Old School" of CRM, the typical answer would be poor communication can lead to an accident and the flight attendants are the last bastion of safety. From the "New School" of CRM the lesson learned would be that "It is alright to cancel the flight. It is the company's responsibility to put up the passengers, NOT the pilot's."

To put this point across, let me cite the following real world war story from a typical CRM class at one of the major air carriers with a CRM program held up as "Providing the Leading Edge." The class opens with a series of pictures showing aircraft destroyed in accidents. Then the following scenario is provided: You are the First Officer on the flight and during the originating flight pre-flight you discover some snow on the wings. You report this to the captain who tells you not to worry about it because it will blow off the wing. You attempt to express that you are uncomfortable with that but the class facilitator keeps telling you that the captain is not listening. The goal of the class is obviously to see how many ways you can tactfully challenge the captain's decision. After a couple dozen attempts to convince the captain his decision is incorrect, one of your peers says "I'm taking my flight bag and leaving the flight deck!" The facilitator freaks out and begins babbling "No, No, you can't do that. You have to keep attempting to convince the captain that his decision is incorrect." That attitude and curriculum have been the foundation for CRM classes all over the world. That approach is NOT always realistic nor always practical; sometimes you just have to say NO and that's what separates the new CRM from the old.

Material for CRM classes should be addressing real world issues and not just something that some committee made up of members from various working groups thought up during a working lunch paid for by the company. Keep the focus on flying and flying issues. GAIN, ASAP, Line Pilot Reports, FAA violations, accident and incident reports, and the ASRS reporting system are all excellent starting points for poignant discussions. Analysis of what was done wrong and right and what might be done differently in the future provides an excellent forum for pilots with tens of thousands oaf of accident free flying hours to share all the tricks of the trade which they have learned over the years to stay out of trouble. Facilitators should have a definite theme for the class and keep the focus of the discussion on that theme. The industry has failed abysmally to take full advantage of this wonderful resource of line pilot experience. Furthermore, each new technological advancement (GPS approaches) or new procedure (Reduced Vertical Separation) brings more challenges to the line pilot and these need to be addressed right along with all the old standard threats of CFIT and running out of gas. Every day that these issues are not addressed is one day closer to the next accident.

HOW HIGH IS HIGH ENOUGH? QUANTIFYING THE IMPACT OF AIR TRAFFIC CONTROL TOWER OBSERVATION HEIGHT ON DISTANCE PERCEPTION

[1]**William K. Krebs**, **Glen Hewitt**
Federal Aviation Administration, Washington, D.C.

Steven R. Murrill
Army Research Laboratory, Adelphi, Maryland

Ronald G. Driggers
US Army RDECOM CERDEC NVESD, Ft. Belvoir, Virginia

Each year the Federal Aviation Administration (FAA) builds approximately seven air traffic control towers in the national airspace system. Each airport has unique surface and airspace characteristics, but all airports must determine the location and height of the new air traffic control tower (ATCT). These two factors impact cost and safety; therefore the FAA must develop a quantitative means in measuring what improvement in ATCT visibility can be gained by increasing tower height at different locations on the airport surface. Two metrics were developed (Object Discrimination, Line of Sight Angle of Incidence) to assess the impact of tower height on distance perception. The two metrics are robust and easy to use to assess the impact of tower height on air traffic control tower specialist distance perception.

Introduction

"The air traffic control tower siting process must take into consideration criteria relating to the safety of air traffic operations for each site. The optimum height and location is the result of balancing many requirements and considerations, based on the current approved Airport Layout Plan (ALP). The goal of this process is to maximize operational performance and safety when siting an ATCT. (6480.xx, page 3)".

A Federal Aviation Administration employee requested assistance in determining a proposed tower height. The employee's request stated:

> *"I've been asked to justify a certain height at a new tower. I've tried to explain to the Terminal Business folks that this place needs a taller tower because of line of sight problems, heat wave distortion, night time glare from lighting that surrounds the airport, and a parallax type of problem when watching aircraft approaching the airport for landing on closely spaced parallel runways. (quote from an FAA employee, 2004)"*

The Federal Aviation Administration William J. Hughes Technical Center Airway Facilities Tower Integration Laboratory (AFTIL) tower cab simulation enables design engineers and air traffic control tower specialists to assess the impact of a proposed tower height and location. The AFTIL can simulate real-world scenes to assess the physical attributes of the tower cab relative to the airport surface and how they may affect visibility. Such attributes include cab orientation, tower look-down angle, look across line-of-site, mullions, look-up angle for missed approaches, movement and non-movement areas, and unobstructed views. The diverse capabilities of the AFTIL entail tradeoffs. For example, to depict a real-world scene in a 360^0 tower cab simulation, the spatial resolution of the generated scene is sacrificed due to amount of computer processing required to generate a scene. In the normal mode, the AFTIL image generated scene is equivalent to 20/80 visual acuity which is more than sufficient to address most of the tower siting criteria. However, the AFTIL can not address the impact of tower height on an air traffic control tower specialist's detection of a distant object.

The objective of this study was to develop, test, and validate a set of human performance metrics to assess the impact of tower height on air traffic control tower specialist distance perception. The human factors metrics as well as the AFTIL simulation will be used to site a tower at an airport.

[1] To whom correspondence should be addressed. Current address: Federal Aviation Administration, 800 Independence Avenue, ATO-P R&D, Room 907A, Washington D.C. 20591 or william.krebs@faa.gov, 202-267-8758.

Methods

Object Discrimination Analysis

Question: What improvement in detecting or recognizing a distant object can be gained by increasing tower height or decreasing tower distance from the object?

The overall objective of this metric is to provide the FAA with a user-friendly software tool that provides quantitative information on the impact of ATCT height on aircraft visibility. The tool includes drop-down windows for user input as well as graphical chart windows for results output. The primary output of this tool is probability-of-discrimination (detection and recognition) curves as a function of observation range and tower height. The tool draws from four well-developed and empirically-validated functions and models: The U.S. Army Night Vision Laboratory's Standard Target Transfer Probability Function (using modified Johnson's discrimination criteria), Barton's model for the human eye's Contrast Transfer Function, Kopeika's atmospheric (optical) turbulence modulation transfer function, and Tatarski's atmospheric-index-structure-parameter height-scaling model. In addition, the algorithms and routines include two enhanced-accuracy features that account for: the impact of turbulence on a downward-slanting optical path, and the effect of distance between the point of optical path integration and the observer (the "shower curtain" effect).

Model Assumptions: The model assumes that

(a) Detection is defined as the ability to notice the presence of an object on the airport surface without regard to the class, type, or model (e.g., an object such as an aircraft or vehicle). The observer knows something is present but cannot recognize or identify the object.

(b) Recognition is defined as the ability to discriminate a class of objects (e.g., a class of aircraft such as single engine general aviation aircraft).

(c) The object (aircraft or vehicle) size is taken to be the square root of the frontal or side cross-sectional area of the object (e.g., wing span x height).

(d) Modified Johnson's criteria is used for the number of optical cycles required for a 50% probability of success in object discrimination (N50).

(e) All observations are made with the unaided eye.

(f) The observer is assumed to be at the specified tower height while all objects (e.g., aircraft, vehicles) are taken to be at the ~ 3 ft (1 m) height.

To account for the impact of atmospheric (optical) turbulence on the downward-slanting optical path, an average/effective refractive-index-structure-parameter *scaling factor* was calculated. This *scaling factor* was derived by taking the line integral of the Tatarski height scaling equation over the downward-slanting optical path.

Object Discrimination Tool: The tool (figure 1) can be found at http://www.hf.faa.gov/visibility.

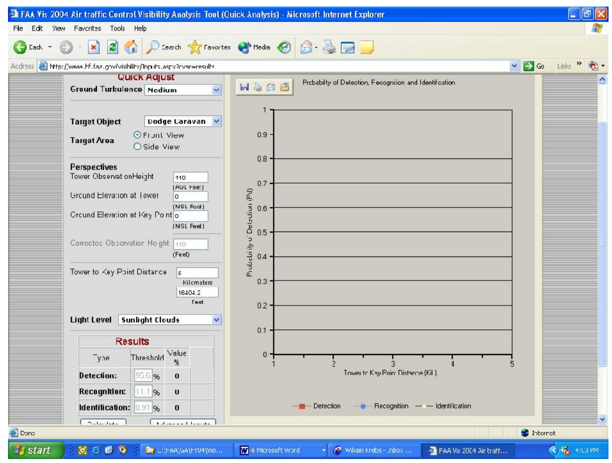


Figure 1. *Object discrimination tool graphical user interface. Users enter tower height and distance to calculate air traffic control tower specialists' detection and recognition of an airport surface object.*

Procedure: From the graphical user interface select an object, specify tower height and key point distance, specify ground turbulence, and specify the outside illumination level. Key point distance is defined as the distance between an observer in the air traffic control tower and object of interest on the on the airport surface.

Results: Probability of detection and recognition values were calculated for one hundred and ninety five grade seven or greater air traffic control towers in the national airspace system. Key point was defined as the most distant runway threshold from the air traffic control tower for each airport. The object selected was a front-view of a Dodge Caravan minivan set at 33% contrast. Illumination was specified as sunlight clouds and ground turbulence was dependent upon geographical location.

Based on the 195 air traffic control tower sample, criterion was set at 1½ standard deviations below the sample mean (i.e., better than 6.7% of the sample) which is equivalent to 95.5% for detection and 11.5% for recognition (table 1).

Observation Capability Requirements	Observation Description	Front View Probability Criteria Minimum
Detection	Ability to notice the presence of an object on the airport surface without regard to the class, type, or model (e.g., an object such as an aircraft or vehicle). The observer knows something is present but cannot recognize or identify the object.	95.5%
Recognition	Ability to discriminate a class of objects (e.g., a class of aircraft such as single engine general aviation aircraft).	11.5%

Table 1. *Probability of discrimination detection and recognition criterion values based on one hundred and ninety five level seven or greater air traffic control towers in the national airspace system.*

Line of Sight Angle of Incidence Analysis

Question: What improvement in the controller's viewing perspective can be gained by increasing the observer's line of sight angle of incidence to the airport surface at key distance points?

Observers: Twelve tower-rated air traffic control specialists, age 26-59 years, were recruited from four different tower airport facilities. Average air traffic control tower experience was 17.4 years. All observers had normal or corrected-to-normal visual acuity, and had normal color vision. All observers granted informed consent prior to participation. All observers were naïve to the experimental hypothesis.

Apparatus: Federal Aviation Administration William J. Hughes Technical Center Airway Facilities Tower Integration Laboratory's (AFTIL) nine Quantum 3D "Alchemy" image generators (IGs) drove nine, six-foot vertical by eight-foot horizontal rear-projection screens arranged in a 360^0 circular pattern to simulate an air traffic control tower cab environment. The diameter of the simulation floor plan is 24'. Each rear-projector, Epson "PowerLight" model 9100, had a pixel resolution set at 1280 (horizontal) by 1024 (vertical) pixels with a field-of-view of approximately 20^0 (horizontal) by 15^0 (vertical). To increase resolution of the visual simulation, three of the nine rear-projection screens were used in the test. Observers were positioned 24' from the most distant screen thereby allowing a resolution of 64 pixels per degree. The base of the screens is approximately 30 inches from the floor to allow an average standing observer's eye-height to be centered on the screen. Software used to model the simulation were AutoCad, MultiGen-Paradigm, PhotoShop, and other graphic simulator tools to generate vehicle ground and air routes for the airport. Frame rate was fixed at 30 frames/second.

Airport Display: The AFTIL tower simulation displayed a realistic depiction of an airport surface using panoramic photographs and computer graphics (figure 2). The visual simulation contained terrain features, hangers, terminals, runways, taxiways, as well as dynamic surface and airborne aircraft and other ground surface vehicles.

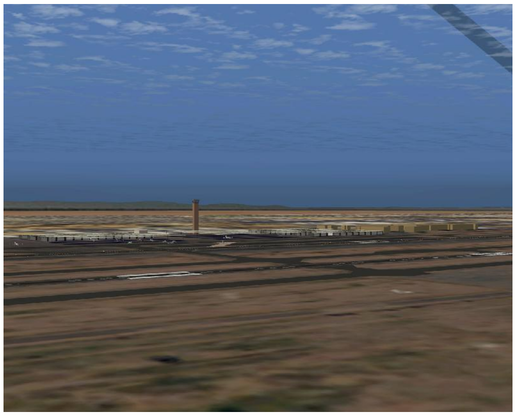


Figure 2. *Simulated air traffic control tower scene generated by the Federal Aviation Administration William J. Hughes Technical Center Airway Facilities Tower Integration Laboratory.*

Eight ATCT simulations were created: Cahokia/Saint Louis Downtown (CPS), Fort Wayne International (FWA), New York/La Guardia (LGA), Memphis International (MEM), Morriston Muni (MMU), Minneapolis-Saint Paul International (MSP), Oshkosh/Wittman Regional (OSH), and Richmond International (RIC). At each airport, a critical key point was selected. Observers were informed on the location of the key point. All simulations were displayed during day illumination.

Procedure: The observer was exposed to fifty experimental dynamic scenes: five of eight ATCT simulations and ten tower observation heights. In each trial, observers performed common air traffic control tower visual tasks at different tower heights. The observer's task was to visually scan a designated distant "key point" on an airport surface and rate the ability to: (1) distinguish boundaries of the movement areas, and (2) identify position of target at the airport's key point. The distant "key point" was an MD-80 located on the airport surface. Prior to entering the tower cab simulation, the experimenter familiarized the observer to a 6-point Likert rating scale and the response criteria for each question. At the beginning of each block of trials, observers were afforded several minutes to familiarize themselves with the airport layout and location of the distant key point. At the completion of the familiarization, the observer's eyes were occluded and the first experimental tower height was selected. The experimenter then instructed the observer to open his or her eyes and respond to both questions. Within each block of trials, tower height was randomly assigned without replacement. At the completion of the tenth tower height, the next ATCT scene was presented and the same procedure was repeated. ATCT scene order was randomly assigned across observers. Reaction time was not recorded.

Results: Calculate the height of the observer in the tower according to the formula:

$$H_O = (H_C - (P_E - T_E)),$$

where, H_O is height of observer; H_C is controller eye height; P_E is ground elevation of key point Above Mean Sea Level; T_E is ground elevation of tower Above Mean Sea Level. Controller eye height is defined as five feet above cab floor height.

Compute the Line of Sight angle at which the observer's view intersects with the airport surface at the key point.

Line of Sight angle = ArcTan (height of observer/distance between key point and tower)

Based on the responses of twelve observers and several other air traffic tower controller specialists, the minimum level of performance for question 1 (*How well can you distinguish boundaries of the movement areas?*) was response 2 (*Can discriminate boundaries of most of runways and taxiways; but provides no distance information*). Figure 3 illustrates observers' proportion of "yes" responses for response of 2 or greater. All observers reported a response of 2 or greater when towers line of sight angle of incidence was 1.5 degrees or greater. Converting the proportion of "yes" responses for response 2 or greater to Z scores, and then fitting a linear line showed that 50% of the observers reported 0.481 degrees as the preferred line of sight angle of incidence (figure 4).

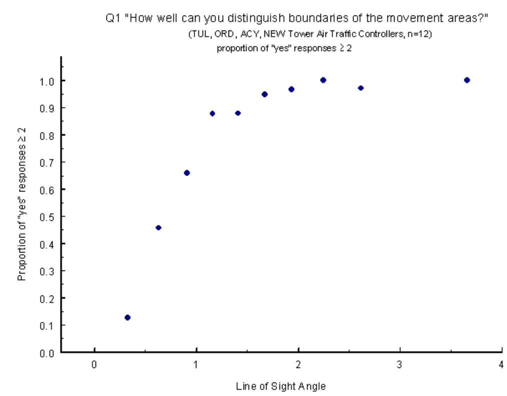


Figure 3. *Illustrates observers' proportion of "yes" responses for response of 2 or greater for question 1: "How well can you distinguish boundaries of the movement areas?"*

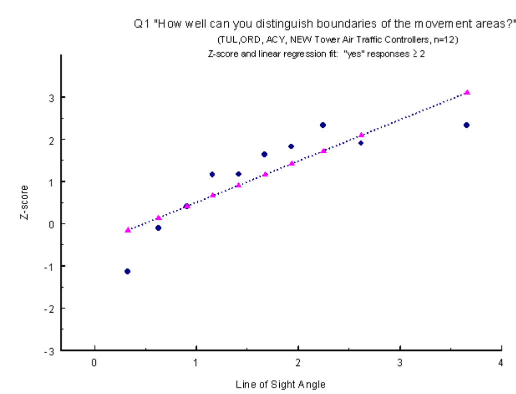


Figure 4. *Converting the proportion of "yes" responses for response 2 or greater to Z scores, and then fitting a linear line showed that 50% of the observers reported 0.481 degrees as the preferred line of sight angle of incidence.*

For question 2 (*How well can you identify the position of an object relative to the airport's key point?*), the minimum acceptable response was 3 (*Able to determine that object position is in general vicinity of key point, but unable to estimate distances of object within movement area*). Figure 5 and 6 illustrate observers' responses for a response of 3 or greater and linear fit to Z scores, respectively. Fifty percent of the observers reported 0.799 degrees as the preferred line of sight angle of incidence (figure 6).

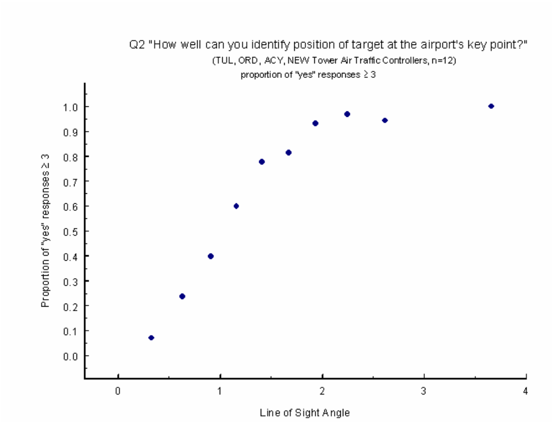


Figure 5. *Illustrates observers' proportion of "yes" responses for response of 3 or greater for question 2: "How well can you identify the position of an object relative to the airport's key point?"*

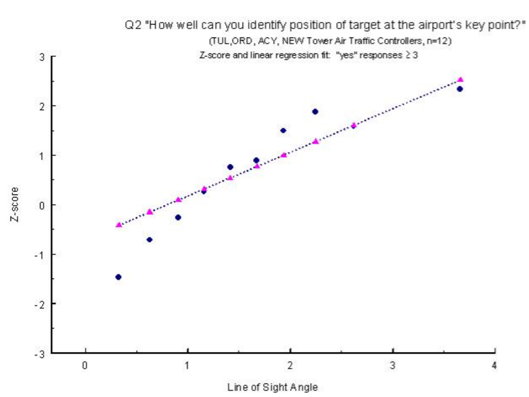


Figure 6. *Observers reported 0.799 degrees as the preferred line of sight angle of incidence for a response of 3 or greater.*

The minimum line of sight angle of incidence is set at 0.799. The higher value was selected because question 2 was reported as the more important task of an air traffic control tower specialist.

Conclusions

The analyses performed may assist air traffic requirements in determining future air traffic control tower heights and location. To assist the decision team, the analyses could be plotted to illustrate percent improvement of air traffic control tower specialists' recognition of an aircraft by tower height expressed in dollars per linear foot. Of course, there are many factors that determine tower height and location but the analyses described above may provide air traffic requirements additional quantitative data to assist in their decision.

References

Federal Aviation Administration (November 10, 1972). Department of Transportation, Federal Aviation Administration Publication Order #6480.4, Airport Traffic Control Tower Siting Criteria.

Federal Aviation Administration (in preparation). Department of Transportation, Federal Aviation Administration Publication Order #6480.xx, Airport Traffic Control Tower Siting Process.

Acknowledgements

Sponsored by Federal Aviation Administration, Human Factors Research and Engineering Division. Also, sponsored by Bill Kaplan, Federal Aviation Administration, Terminal Facilities.

Thanks to Scott Iwamoto, Kathy Peterson, Dave Cloutier for their never ending support in allowing us to participate in the air traffic control tower siting process, and a special thanks to Bill Vaughan for providing access to the William J. Hughes Technical Center Airway Facilities Tower Integration Laboratory. We also thank John Aschenbach, Rodman Bourne, Roger Bawgus, and Bernie Garbowski in their assistance with the simulation data collection, and Rob Heckart and Doug Norvell for placing the tower visibility tool on the website.

The views expressed in the article are those of the authors, and do not reflect the official position of the Department of Transportation, Department of the Army, the Department of Defense, or the United States Government.

ASSESSING HUMAN FACTORS RISKS IN AIR TRAFFIC MANAGEMENT RESEARCH

Paul Krois
Federal Aviation Administration
Washington, DC

Jacqueline Rehmann
Federal Aviation Administration Technical Center
Atlantic City, New Jersey

This paper describes guidance for assessing human factors risks of research capabilities. The risk scales in the present study were used to assess concepts of the FAA Target System Description, part of the National Airspace System architecture. Results help identify areas where human factors analyses are needed and strengthen the business case for human factors assessments. Repeated high risk ratings provide R&D managers with additional information about whether or not to proceed with specific capabilities.

Introduction

The Federal Aviation Administration (FAA) System Engineering Manual defines risk as "a future event or situation with a realistic (non-zero nor 100 per cent) likelihood of occurring and an *unfavorable consequence/impact* to the successful accomplishment of the well-defined program goals if it occurs" (2004a). Relative to the FAA's definition, a consistent and more quantitative approach to human factors risk management is needed for application to research and development (R&D) on air traffic management. This paper describes the guidance developed for assessing human factors risks of R&D capabilities.

Previous research (Krois, Mogford, and Rehmann, 2003) posed that select human factors study areas should be incrementally addressed as an R&D prototype matures. That research also validated the need to consider all human factors study areas defined in the FAA Job Aid (2003a) by the time a research prototype is tested in a laboratory with representative users. These findings provided programmatic guidance for early, realistic assessments of innovative concepts and emerging technologies for use in an integrated National Airspace System (NAS). The research described in this paper poses that from a risk standpoint, the same human factors study areas could be assessed to better gauge and understand potential program impacts of research capabilities.

The human factors study areas pose potential risks in relation to the conceptual service improvements comprising the FAA Target System Description (TSD), which is part of the NAS architecture (FAA, 2004b). Researchers need help to strengthen the business case for human factors studies that resolve risks. This includes assessments that use complex human in the loop (HITL) simulation. Resolution of high risk areas provide R&D managers with additional information about whether or not to continue with development of specific capabilities.

Technical Readiness Levels

One approach for assessing the maturation of capabilities from laboratory to field implementation is the Technology Readiness Level (TRLs) model. TRLs are described in the FAA/NASA Integrated Plan for Air Traffic Management Research and Technology Development, Version 7.0 (FAA, 2003b). The TRL model intends an orderly transfer process from R&D to deployment and has been used by NASA and FAA to define transitions between stages of R&D leading to NAS implementation. FAA recognizes the need to manage more efficiently the transition of R&D capabilities, promote understanding of human factors risks, and make informed program decisions. The TRL model helps researchers from different organizations coordinate objectives and outputs and assess research maturation.

TRLs 1-6 pertain to concept exploration and concept development phases and consist of the following. TRL 1 Basic Principles Observed/Reported is the stage at which a capability is initially identified and described. In TRL 2 Technology Concept and/or Application Formulated a research plan is developed that defines the technical solution to the deficiency identified in TRL 1. This plan identifies activities, schedule, and resources necessary to address issues with the capability. In TRL 3 Analytical/ Experimental Critical Functions or Characteristic Proof of Concept, a conceptual prototype of the capability is developed and initial requirements are defined. The use of metrics to assess benefits should show an improvement over the baseline. TRL 4

Component or Integrated Components Tested in a Laboratory Environment is the stage where a research prototype of the capability is developed and evaluated by representative users. The laboratory real time simulation environment is at a higher fidelity level than at TRL 3. In TRL 5 Components/Subsystems Verified in a Relevant Environment, a pre-development prototype is prepared and evaluated. The evaluation environment should be at a high fidelity. The FAA assumes "ownership" of the operational concept and initiates activities to transition the capability to the acquisition product team. In TRL 6 System Demonstrated/ Verified in a Relevant Environment, an operational demonstration of the pre-production prototype system is conducted in an FAA field facility.

Human-System Interaction Research

Human factors research addresses human-system interaction (HSI) as R&D capabilities mature. This is shown through technical reviews of ATC modernization improvements planned in different operational environments, lower-fidelity assessments such as cognitive walkthroughs, and complex HITL simulations including integrated air-ground simulations.

As one example, "allocation of function" is one of the earliest human factors considerations effecting successful HSI. Inadequate consideration of allocation of function was evidenced in an operational evaluation of a controller decision support tool called passive Final Approach Spacing Tool (pFAST). In the evaluation, controllers found some advisories problematic but were instructed to use them anyway unless the non-use was approved by supervisors. Unfortunately, pFAST was found to work best when all controllers followed all advisories, i.e., pFAST performance degraded when advisories were not used. With pFAST the flexibility in decision making that controllers routinely exercised became constrained (Cardosi, 2002).

Trends suggest that information management demands associated with effective HSI are changing from being based on procedural requirements to controller workload impacts. Automation poses the risk of tunneling the attention resources of the controller and molding the application of those resources so as to be opaque to unforeseen and subtle events and incidents. Experience with another controller decision support tool called URET indicates that controllers make decreased use of trial planning as workload increases, and that use of the traffic management advisor (TMA) metering list decreases when workload is high. Such "automation shedding" moves the controller into a manual control mode that poses a hysteresis effect for returning to use of automated tools, e.g., the controller has to restart building a new mental picture of the traffic situation and not reverting to a previous mental state.

Human Factors Study Areas

Human factors study areas used to assess research risks are taken directly from the FAA Human Factors Job Aid and consist of the following.

- Allocation of Function: Assigning those roles/functions/tasks for which the human or equipment performs better while enabling the human to maintain awareness of the operational situation.
- Anthropometrics and Biomechanics: Accommodating the physical attributes of its user population (e.g., from the 1st through 99th percentile levels).
- Computer-Human Interaction (CHI): Employing effective and consistent user dialogues, interfaces, and procedures across system functions.
- Communications and Teamwork: Applying system design considerations to enhance required user communications and teamwork.
- Culture: Addressing the organizational and sociological environment into which any change, including new technologies and procedures, will be introduced.
- Displays and Controls: Designing and arranging displays and controls to be consistent with the operator's and maintainer's tasks and actions.
- Documentation: Preparing user documentation and technical manuals in a suitable format of information presentation, at the appropriate reading level, and with the required degree of technical sophistication and clarity.
- Environment: Accommodating environmental factors (including extremes) to which the system will be subjected and understanding the associated effects on human-system performance.
- Functional Design: Applying human-centered design for usability and compatibility with operational and maintenance concepts.
- Human Error: Examining design and contextual conditions (including supervisory and organizational influences) as causal factors contributing to human error, and consideration of objectives for error tolerance, error prevention, and error correction/recovery.
- Information Presentation: Enhancing operator and maintainer performance through the use of effective and consistent labels, symbols, colors, terms, acronyms, abbreviations, formats, and data fields.
- Information Requirements: Ensuring the availability and usability of information needed by the operator

and maintainer for a specific task when it is needed, and in a form that is directly usable.
- Input/Output (I/O) Devices: Selecting I/O methods and devices that allow operators or maintainers to perform tasks, especially critical tasks, quickly and accurately.
- Knowledge, Skills and Abilities (KSAs): Measuring the KSAs required to perform job-related tasks, and determining appropriate selection requirements for users.
- Operational Suitability: Ensuring that the system appropriately supports the user in performing intended functions while maintaining interoperability and consistency with other system elements or support systems.
- Procedures: Designing operation and maintenance procedures for simplicity, consistency, and ease of use.
- Safety and Health: Preventing/reducing operator and maintainer exposure to safety and health hazards.
- Situational Awareness: Enabling operators or maintainers to perceive and understand elements of the current situation, and project them to future operational situations.
- Special Skills and Tools: Minimizing the need for special or unique operator or maintainer skills, abilities, tools, or characteristics.
- Staffing: Accommodating constraints and efficiencies for staffing levels and organizational structures.
- Training: Applying methods to enhance operator or maintainer acquisition of the knowledge and skills needed to interface with the system, and designing that system so that these skills are easily learned and retained.
- Visual/Auditory Alerts: Designing visual and auditory alerts (including error messages) to invoke the necessary operator and maintainer response.
- Workload: Assessing the net demands or impacts upon the physical, cognitive, and decision-making resources of an operator or maintainer using objective and subjective performance measures.
- Work Space: Designing adequate work space for personnel and their tools or equipment, and providing sufficient space for the movements and actions that personnel perform during operational and maintenance tasks under normal, adverse, and emergency conditions.

FAA Target System Description (TSD)

FAA is modernizing the NAS, nominally to achieve the Joint Concept of Operations (CONOPS) as described by the RTCA (2002). The TSD describes improvements to NAS service capabilities expected by 2015. It includes descriptions of systems to be implemented, services provided, and operational capabilities that will be achieved. Together the NAS Architecture and the TSD provide a roadmap to guide the evolution of automation, surveillance, navigation, and communication systems to ensure NAS modernization is achieved.

The TSD, like the NAS architecture, is organized into 9 service areas including flight planning; separation assurance; tactical traffic flow; strategic flow; advisory services; emergency and alerting; navigation; airspace management; and infrastructure and information management. These are further decomposed into service improvement areas identified later in this paper.

Key HSI characteristics of TSD operational improvements consist of the following.
- A system wide information management system (SWIM) will serve as the central depository for all NAS information.
- Wide-spread, real-time distribution of NAS data.
- A standard automation platform (SAP) will be used by both terminal and en route controllers.
- Decision support systems (DSS) and intelligent agents will be common.
- Maximum use of digital communications.
- Maximum use of ADS-B for surveillance.
- Traffic managed gate to gate.
- Integrated ATM/CNS provides seamless airspace (Surface, Terminal, En Route and Ocean).
- Flexible airspace to match the dynamics of demand.
- Three mile separation used throughout the airspace.
- Pilots participate in managing aircraft separation.
- Airborne and ground conflict alerting.
- Auto-negotiations to develop flight profiles.

Methodology

We conducted an analysis to assess the viability of using specially developed scales for the 24 human factors study areas in helping to gauge risks associated with the TSDs. The resultant risk ratings reflect an average response as determined by human factors subject matter experts who participated in the analysis.

Human Factors Risk Scales

Risk scales used a five point scale based upon similar scales in the FAA System Engineering Manual (FAA, 2004a) and a safety risk assessment approach developed for military product improvements (Naylor, 2000). A low risk associated with the human factors study area is assigned a numerical value of 1; accordingly, if there is a minor risk, the value assigned is 2; a moderate risk is assigned a value of 3; a significant risk was assigned a value of 4; and a high risk associated was assigned a value of 5. A sample risk scale is as follows.

Allocation of Function: System design reflects assignment of operational roles, functions, tasks to humans or equipment while maintaining the human's awareness of the operational situation.
Low: Allocation of function of the proposed R&D capability does not change the current roles, functions, and tasks presently assigned to humans or equipment nor changes the operator's situation awareness.
Minor: Integration of the R&D capability into the present work environment may result in limited changes to current roles, functions, and tasks presently assigned to humans or equipment and may slightly alter the operator's situation awareness.
Moderate: Integration of the R&D capability into the present work environment alters current roles, functions, and tasks presently assigned to humans or equipment, and impacts the operator's situation awareness.
Significant: Integration of the R&D capability into the present work environment significantly alters current roles, functions, and tasks presently assigned to humans or equipment, and significantly impacts the operator's situation awareness.
High: Allocation of function of the proposed R&D capability alters completely the current roles, functions, and tasks presently assigned to humans or equipment and changes completely the operator's situation awareness such that how an operator's 'mental picture' is formed no longer exists.

Results

It should be noted that the data and its analysis are notional and illustrate one approach to help identify human factors risks. Average scores for the 24 human factors study areas are shown in Figure 1, based on ratings across the 19 service improvement areas. Results showed the topmost risks as Allocation of Function, Communication and Teamwork, Procedures, Information Requirements, Workload, Human Error, Culture, Information Presentation, and Situation Awareness.

The average scores for the 19 service improvement areas are shown in Figure 2. Results suggest a clustering of a small number of areas posing the highest total human factors risk consisting of Aircraft-Aircraft Separation and Flight Data Management. Another cluster with a large number of areas of high risk included Airspace Management, Traffic Advisories, Surface Separation, Monitoring and Maintenance, Flight Plan Support, Airborne Synchronization, Airspace Design, Alerting Support, Aircraft-Terrain Separation, Surface Synchronization, and Weather Advisories.

In the course of assessing the service improvement areas, questions, issues, and potential interdependencies influencing human factors risk were identified to help clarify the basis for ratings.

Several questions/concerns raised by human factors experts help clarify ratings of 4 or 5. For example, for the aircraft/aircraft separation service improvement, human factors questions include the following considerations:
- Communications and Teamwork: What is the impact associated with changing roles and responsibilities and how will separation authority transition between controller and pilot?
- Culture: What is the impact from divergence of operating norms and business cultures between FAA air traffic controllers and multiple unique airlines?
- Functional Design: How compatible are the alerting logic algorithms among airborne and ground systems and what will be the impact on pilot and controller decision making?
- Human Error: What is the potential for pilot and air traffic controller task performance error during critical operations?
- Information Presentation: How will information be displayed?
- Information Requirements: What information is needed?
- Procedures: How will tasks and procedures change?
- Situation Awareness: What are the impacts on controllers' situation awareness?
- Special Skills and Tools: What human performance considerations are associated with reduced separation?
- Visual/Auditory Alerts: Will air and ground trajectory models and conflict prediction algorithms be integrated?
- Workload: What are the workload impacts on air traffic controllers of using new automation and decision support tools for reduced separation tasks?

Discussion

Assessment of human factors risks shows that higher risk ratings occur in relation to the degree to which roles and responsibilities of controllers and pilots change. Our understanding of this relationship should improve as service improvements are further defined and implemented. For example, the role of the controller will drive operational requirements for information. This information is used in accordance with new procedures that pose changes in workload and communications, as well as the potential for human error.

Previous research yielded guidance that as a research capability matures and progresses from TRL-1 to TRL-6, the set of human factors issue areas increases and becomes more critical (Krois, Mogford & Rehmann, 2003). In contrast, the efficacy of research is proportional to human factors risks being addressed and attenuated as the capability progresses across the TRLs. The argument to graduate a research capability beyond TRL-6 is strengthened by the extent that human factors issues/risks are identified, assessed, and resolved within the context of interoperability with the baseline operational system.

The integration of human factors tools and techniques in the TRL model is important to the smooth transition of capabilities from research concepts to field-ready systems and procedures. FAA has found that while human performance issues may or may not impose constraints on cost and schedule, they will be the limiting factor in achieving system performance, thus impacting successful deployment. In the Standard Terminal Automation Replacement (STARS) program, for example, FAA found that system design and system development efforts can lose sight of the human-system performance impacts on end users (operators, maintainers, and support personnel), especially those related to cognitive tasks. History has taught us that if these elements are closely attended to during system development and prototyping, FAA has estimated saving a few months to up to 18 months in program development time and savings of up to 20% of program costs (FAA STARS Human Factors Evaluation, 1998).

Conclusions

It is important for research to have a risk management strategy in place to help identify mitigation strategies that may be employed. The transition from research and operational prototypes to development and fielding is complex. It is not only important to have specific criteria that define a concept's readiness to transition from one state to the next but also to have a means of assessing potential human factors risks especially when the capabilities and technologies entail new roles for controllers and pilots. The risk scales developed for this study are intended to serve as a tool to researchers to help them better understand and quantify the impact of specific human factors risks as research progresses.

Disclaimer - The views expressed are those of the authors and do not represent the FAA.

Acknowledgement - An analysis of the TSDs was previously conducted by the FAA Human Factors Research and Engineering Division.

References

Cardosi, K. (2002) Human Factors Integration Challenges in the Terminal Radar Approach Control (TRACON) Environment. Cambridge, MA: John A. Volpe National Transportation Systems Center.

Federal Aviation Administration (2004a) *FAA System Engineering Manual v3.0*. Washington, DC: U.S. Department of Transportation.

Federal Aviation Administration (2004b). National Airspace System Architecture 5, http://www.nas architecture.faa.gov/cats/faq/tsd/

Federal Aviation Administration (2003a). *Human Factors Job Aid*. Washington, DC: U.S. Department of Transportation.

Federal Aviation Administration and National Aeronautics and Space Administration (2003b). *Integrated Plan for Air Traffic Management Research and Technology Development, Version 7.0, Volume I: Plan Overview and Joint Research Project Descriptions*. Washington, DC: U.S. Department of Transportation.

Federal Aviation Administration. Acquisition Lessons Learned: Solution Implementation, Human Factors Record 99-006, STARS Program, 1998.

Krois, P., Mogford, R., & Rehmann, J. Integrating Human Factors Issues in Technology Readiness Levels for Air Traffic Control Research. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting, 2003.*

Naylor, W.P. (2000) *Assessing Safety and Programmatic Risks of Proposed System Upgrades*. Journal of System Safety, pp 25-32**.**

RTCA (2002). National Airspace System Concept of Operations and Vision for the future of Aviation. Washington, DC.

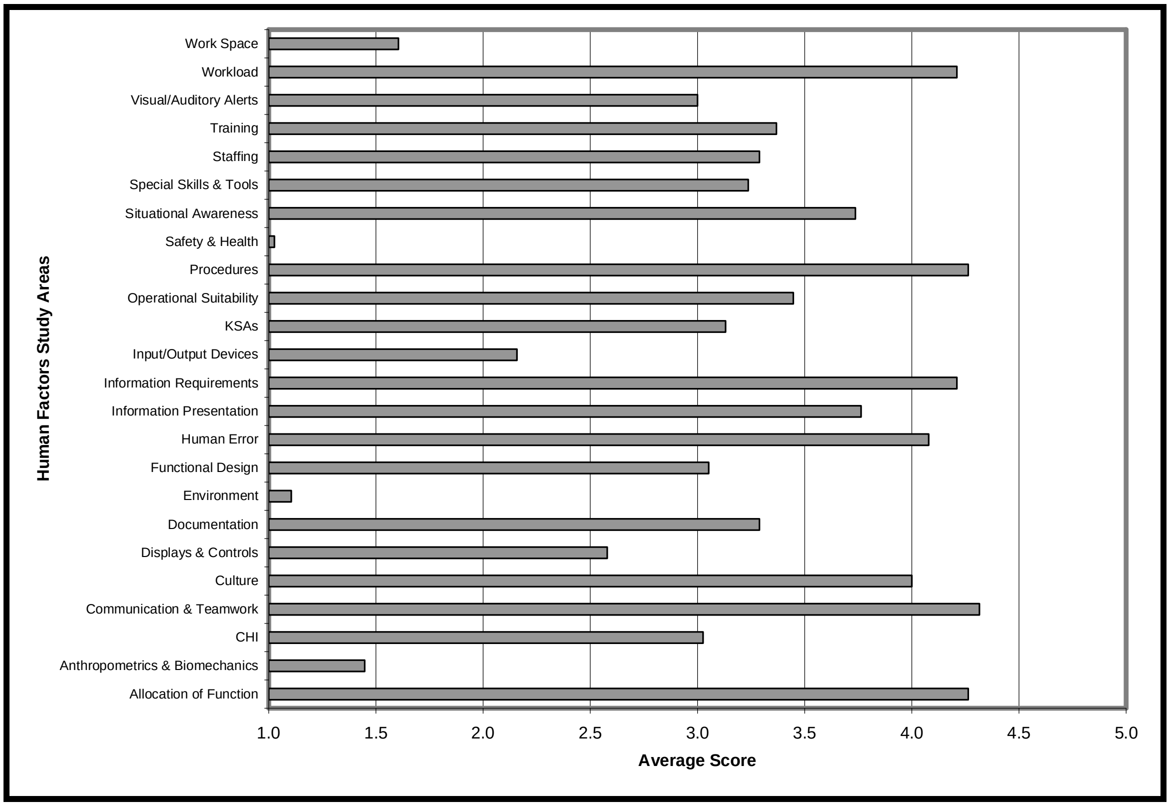


Figure 1. *Average risk ratings for 24 human factors study areas.*

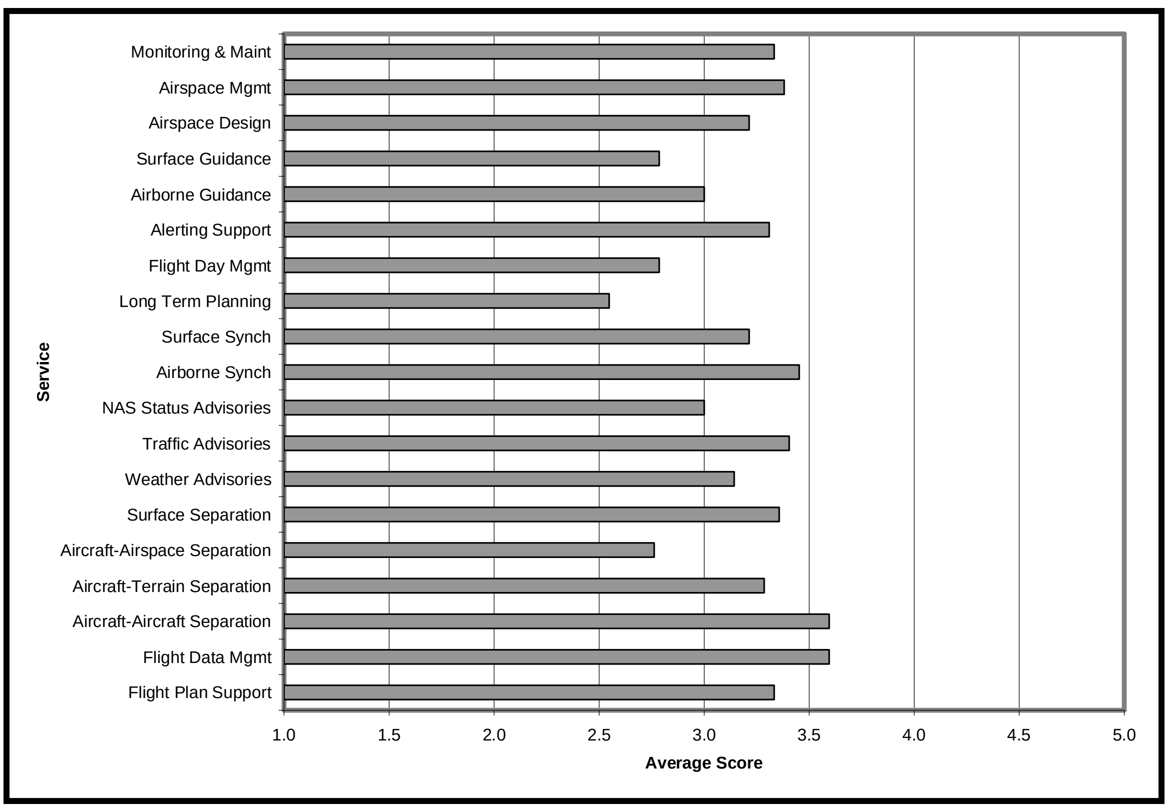


Figure 2. *Average risk ratings for 19 service improvement areas.*

THE CONSIDERATIONS OF A SOUTHEASTERN EUROPEAN AIRLINE AIRCRAFT ENGINEERS REGARDING THE INCENTIVES APPLIED BY THE CORPORATE TOP MANAGEMENT AND THOSE REQUIRED BY THEM

Dr Yannis (John) Lainos
6 Papaflessa st Paleon Faliron-Athens Greece P.C. 17563
Aircraft Technology Department Halkis Technological Educational Institution-Greece
E-mail Lainos@otenet.gr Tel. (+30) 210 98 49 103

In October 2003 we conducted a survey among the aircraft engineers of a southeastern European airline (Airline). The research included four questionnaires with 61 questions. The subjects of these questionnaires were:

a) The incentives that attracted the a/c engineers to their profession.
b) The opinion of the a/c engineers on the incentives applied by the Airline's top management
c) The incentives which the a/c engineers require to be applied by corporate top management.
d) The opinion of the a/c engineers regarding the effects of the recently organisational restruring of the Airline's group

The sample used for this research represented 25% of the total number of the a/c engineers. This percentage is regarded as statistically adequate for an absolutely reliable result.

The construction of the questionnaires and the classification of the answers of the a/c engineers was based on Herzberg's Needs theory. which was adopted at the particularities of the national and labor culture.

Keywords

Airplane: engineers, maintenance unit, culture (national, labor) labor relations, incentives, perception determinants.

The Aim of the Paper

The aim of the present paper is to determine

a) The reasons for the Airline's a/c engineers' assessment of the corporate incentives (applied and required).
b) The impact of the corporate and social environment on shaping these considerations.

The Author's Contribution

The author's contribution is the determination the internal and external Airline corporate environment factors, that contributed to shaping the particular consideration by the Airline's a/c engineers regarding the incentives applied by the top management and those required by them.

Introduction

The southeastern European Airline which is the subject of our case study is a State-owned former flag carrier that has been 100% nationalized since 1975.

The interference of the government, the political parties and union mechanisms in the Airline's management had disastrous results not only on its competitiveness but also on its very survival. (Lainos 1992, Verelis 2004) The crisis was manifested after 1993 when the EU applied the open-skies policy, establishing the air transport liberalization. The state (the Airline's exclusive shareholder) from 1993 to 2004, after six efforts, failed to make the Airline competitive and to privatize it. Finally in December 2003 the flight division of the Airline, as well as that of its regional subsidiary, was spun off and merged with its charter subsidiary company that was re-named. The maintenance unit remained with the old structure although its main client is still the Airline. (Lainos 2003) Our research was conducted during October 2003 with the solidarity of the Airline management and the a/c engineers union. The answers of the respondents in our research were obviously affected by the Airline's group organisational restructuring, which was in process during this period of time.

Our approach regarding the causes of the:

a) particular prioritisation of the incentives which attracted the a/c engineers in their profession and the incentives they require to be applied by the Airline's corporate top management
b) a/c engineers consideration regarding the effects of the Airline's group organisational restructure on the survival of the maintenance unit and on their labor positionis based on the results of the elaboration of the answers which gave the majority of the a/c engineers participants in our research during personal interviews after they had answered the questionnaires.

The construction of the questionnaires and the categorisation of the answers was based on Herzberg's needs theory (Herzberg 1966) adopted at the particularities of the national and labor culture of the Airline a/c engineers.

Brief Theoretical Approach

Herzberg based his dual factors needs theory, (Herzberg, Mausner, Synderman, 1959) on Maslow's five levels Hierarchy of Needs theory (Maslow, 1954). Maslow's needs theory reflect all the needs in a person's life. Maslow argued that the lower level has to be satisfied so that the human will proceed in satisfaction of the next group of needs

Herzberg developed a list of factors that are more closely related to labor environment. These factors fall into two groups. Hygiene or Maintenance factors or Diss*atisfiers and the* Motivators or Satisfiers . (Herzberg 1966, 2003), The hygiene or maintenance factors or disatisfiers are shown in TABLE 1

Table 1. *Hygiene or Maintenance factors or Diss*atisfiers

1) Salary and benefits
2) Personal life
3) Labour conditions
4) Job security
5) Quality of supervision
6) Fellow workers
7) Corporate policies and administrative practices

Herzberg argued that these group of factors do not have motivation power because they will not increase the employees satisfaction within their job. However they will help remove feelings of dissatisfaction. They include a decent salary, acceptable working conditions and the way a company views and treats its employees.

The hygiene or maintenance factors or disatisfiers must be satisfied in the job before the application of motivators-satisfiers. Since hygiene factors are reliably met, the following second set of motivations-satisfiers arises. (TABLE 2) Herzberg argued that these factors are the real motivators.

Table 2. *Motivators-Satisfiers*

1) Status
2) Recognition
3) Promotion
4) Responsibility
5) Development
6) Achievement
7) Job challenge

The Results of Our Research

A) The top 6 incentives that attracted the a/c engineers in this particular a/c Maintenance Unit to their profession

The top 6 incentives that were selected by more than 50% of the a/c engineers participated in our research-are shown in TABLE 3

Table 3. *The top 6 incentives that attracted the a/c engineers of the particular Maintenance Unit to their profession*

1) 73,1% Job stability, certainty, permanency
2) 73,0% The satisfaction derived from the Airline's and the Maintenance unit vanguard in the field of flight safety
3) 62,0% The wage level
4) 61,4% Satisfaction from the achievements of their labour
5) 57,9% The content and social status of their profession
6) 53,8% The social status of the Airline

Classification of the Answers According to the Herzberg's Hygiene and Motivational factors

The received answers of "A" questionnaire were classified according to Herzberg's labor incentives structure, are shown in TABLE 4

Table 4. *Classification of the a/c engineers answers according to the Herzberg's Hygiene and Motivational factors*

5 ACHIEVEMENT…………..…**4rth 61,4%** RESPONSIBILITY JOB CHALLENGE……...…..….**5th 57,9%**
4 RECOGNITION……………..…..**2nd 73%** ADVANCEMENT STATUS…………**5th 57,9% & 6th 53,8%**
3 CORPORATE POLICIES & ADMINISTRATIVE PRACTICES, FELLOW WORKERS QUALITY OF SUPERVISION
2 JOB SECURITY …………..…**1rst 73,1%** LABOR CONDITIONS
1 PERSONAL LIFE SALARY & BENEFITS………....**3rd 62%**

The Causes for the Particular Prioritization of the Incentives That Attract the Maintenance Unit Engineers to Their Profession

1) 73,1% Job stability, certainty, permanency

We consider that the reasons that guided the majority of the a/c engineers of this southeastern European Airline to select this motive in their first priority are the following:

The National Culture

A basic pillar of the Greek national culture is Uncertainty Avoidance, (From scale 40 to 112 Greeks score 112*)*. This dimension addresses the ease with which cultures cope with novelty, ambiguity and uncertainty. (Hofstede 1980 a, b, 1991; Johnston 1993)

The Labor Culture

Even though the position of the Airline which is the main client of the Maintenance Unit being surveyed, was and still is unstable, the a/c engineers believe that the security of their job is not in danger. The factors which contributed to shaping this particular consideration are:

a) The country's need for at least one airline. Consequently, they consider that one aircraft Maintenance and Repair Unit is necessary.
b) Their awareness of the minimum four years required to educate a B1 and B2 aircraft engineer
c) Their experience from the Maintenance unit of the Belgian flag carrier Sabena, which is still in operation even though the airline Sabena ceased its operation since 2002.
d) The quality of their job. This particular Maintenance unit was awarded a prize from the European Maintenance Management Academy (Auditing Organisation approved by IATA). in 1999
e) The power of the union. One hundred per cent of the personnel is unionized, with a history of strong and long term struggles. Their most famous strikes were against the military junta in 1973 and the 70- day strike against the conservative government in 1980.
f) The consideration that their jobs are not in danger is promoted mainly by the aircraft engineers' union.

2) The incentives prioritised in two and four

Priority of TABLE 3 (Satisfaction with the Airline's and its maintenance unit world flight safety record, and satisfaction from the achievement of the a/c engineers labor) have the same content, which is the a/c engineers satisfaction with their job's achievement. This consideration is an international incentive necessary for aircraft engineers to compensate for the stress resulting from their professional responsibility to maintain the aircraft airworthiness that protects passengers' lives. The motto of the International Union of the Aircraft Engineers is "We keep them flying safely". The a/c engineers of this particular Airline have developed this consideration as a corporate culture due to the flight safety record of their company and its maintenance unit.

3) Incentive number three The wage level (62,0%) is related to the decline of the market value of their salaries over the last 15 years.

4) Incentives number five and six **(**The content of the profession and its social status and The social status of the Airline) besides reflecting an existing social consideration, is related to the number two consideration of the a/c engineers.

B) The consideration of the a/c engineers regarding the effectiveness and efficiency of the incentives applied by corporate management

Some 80,2% consider that the Airline management does not apply any incentives for them.

The Causes of the Particular Consideration

The elaboration of the answers, of the a/c engineers participants in our research, during their personal interviews, concluded that the causes of this approach are considered to be the following:

1) The inefficient and ineffective policy applied by the government, the Airline's exclusive shareholder, and the corporate top management appointed by it, regarding the measures applied for survival and the growth perspectives of the Airline and its maintenance unit
2) The non meritocracy criteria of promotion
3) The decrease buying power of their wages over the past fifteen years

C) The top 7 incentives which the Airline's aircraft engineers want to be applied by the corporate top management

The top seven incentives out of the 13, included in this questionnaire, which were selected by more than 50% of the a/c engineers participants in our research are showed in TABLE 5

Classification of the Answers According to the Herzberg's Hygiene and Motivational factors

The received answers of the "C" questionnaire were classified according to Herzberg's labor incentives structure, as TABLE 6

A simple glance at the answers in the questionnaire regarding "The top 6 incentives that attract the personnel of this particular airplane Maintenance Unit to their profession" and the questionnaire answers regarding "the consideration of the a/c engineers regarding the effectiveness and efficiency of the incentives applied by corporate management" indicates an obvious incompatibility. The incentives which attract the a/c engineers at their profession are focused mainly on the Herzberg's satisfiers while the required to be applied by the corporate top management are focused mainly on Herzberg's disatisfiers. We consider that the cause is the incopatibility between the consideration of the a/c engineers regarding their current professional and social status and the objective reality

Table 5. *The top 7 incentives which the Airline's aircraft engineers want to be applied by the corporate top management.*

1) 82% Improvement of their wages level
2) 69,7% Improvement of their working conditions
3) 65,6% Improvement of their training level
4) 59,8% The interest of the Airline's top management in solving the a/c engineers professional problems
5) 59,0% Application of measures for recovery of the Airline.
6) 58,2% The Airline's top management should permit a/c engineers to submit their suggestions for improving the operation of the Airline and the Maintenance Unit and to take into account and to implement these proposals
7) 52,4% The Airline's top management should offer bonuses (extra wages, additional days off, paid vacations, etc)

The Causes of these Considerations

The reasons for the incentives that the a/c engineers require to be applied by corporate management and their prioritisation are:

1) Regarding the requirement for improving their wage level: Twelve years ago (1993), when the government applied for the Airline a "recovery program" approved by the E.U., the purchase power of their wages decreased owing to incompatibility of wage increases and inflation.

2) Regarding the requirement for improving their working conditions:

a) Over the past seven years three aircraft engineers died most probably of cancer owing to the benzolium they come in contact with.
b) The government's effort to privatise the company meant that the tools of their profession were not updated properly.
c) The a/c engineers consider that their job is not efficiently and effectively organised

3) Regarding the requirement to improve a/c engineers training level: A quick comparison between the a/c engineers' second answer to

Table 6. *Classification of the answers according to the Herzberg's Hygiene and Motivational factors*

questionnaire "A", according to which they are proud of being in the international vanguard of the safety of their Maintenance Unit, and their requirement to improve their training level could generate ambiguities relative to the compatibility of these two answers. How could a Maintenance Unit be in the forefront of flight safety without a high level of aircraft engineer training? But in fact, there is no contradiction. The essence of the a/c engineers' requirement that corporate management improve the training level is mainly economic. The Airline's fleet includes more than three aircraft types. The a/c engineers need more than four years to acquire certificates for these aircraft types. According to their collective agreement, when an a/c engineer acquires the certificate for each aircraft type, his salary increases on a percentage basis. So the deeper essence of this requirement is to accelerate their training in order to acquire certificates for all types of aircraft in the fleet in order to increase their earnings.

4) The number four requirement is related to numbers five and six (TABLE 5). It expresses the disappointment of the a/c engineers with the inability of corporate top management to manage general issues regarding the current operation and growth of the aircraft Maintenance Unit. Thus their union is obliged to become involved with these issues even though they are not included in its typical role. For example:

a) The union presses top management to expand the Maintenance Unit activities internationally to cover an existing demand, in order to increase the corporate income. However top management has not responded to this demand arguing that they have insufficient available qualified person- nel. (Statement of the Airline C.E.O. 2004)) However simultaneously the top management do not hire the proper personnel.

b) Despite the 100% increase of the a/c engineers' (and other specialists') job productivity over the last ten years, the financial results of the Airline showed a deficit. (Lainos 2004)

a) During the government's negotiations in the summer of 2000 to privatise the Airline, the investor stated that he could not accept some contracts regarding aircraft acquisition by the Airline because the price was about 50% higher than the corresponding market prices. (Manos 2001)

The a/c engineers argue that their contribution to corporate survival and development has been canceled by such malfunctions.

4) The number five requirement expresses:

a) the disappointment of a/c engineers at the ten years of ineffective and inefficient efforts by the Airline's top management and the government, as its exclusive shareholder, to ensure the Airline's survival and to give it a developmental perspective. They are unwilling to accept that this failure has not been due to incapability, but is the result of a government plan to scuttle the survival effort and thus to persuade the society of the necessity to privatise the Airline, despite its world record in flight safety that was acquired during its state-owned period. (Greek Parliament 2002, Lainos 2003)

b) the inability of corporate top management and the government, the Airline's exclusive shareholder, to protect it from the mass slander campaign by the press. The aircraft engineers do not want to accept that this slander campaign is tolerated if not coordinated by the Airline's exclusive shareholder and the corporate top management appointed by the shareholder. The aim of the shareholder is to persuade the society of the necessity of privatising the flag carrier. (Lainos 2003)

5) Requirement number six expresses the a/c engineers' conviction of their ability to use their knowledge to contribute to the survival and development of the Airline group. The first determinant of this consideration is based on the a/c engineers' upgraded self confidence due to the international recognition of the superior quality of their work. The second determinant is their union's belief that they collectively possess sufficient professional qualifications to realise this aim effectively. The union has organised successfully in recent years:

a) a conference of the International Association of Aircraft Engineers (AEI 2001) and

b) professional conferences regarding the causes of the Airline's crisis and the requirements for its competitive growth (2004)

6) The number seven incentive (The Airline's top management should offer bonuses (extra wages, additional days off, paid vacations and international travel, etc), supported by 52,4% of the participants, expresses the way in which the a/c engineers require that corporate top management actively recognise their effort of the Maintenance Unit to retain its flight safety record.

D) The opinion of the a/c engineers regarding the effects of the recent organisational restructuring of the Airline's group on their job status and on survival and development of the maintenance unit is shown in Table 7.

Table 7. *The causes of the aforementioned answers*

1) Impact on the a/c engineers job position
a) 74,2% Will be worsen
b) 18,1% Will remain stable
c) 8,2% Will be improved
2) Impact on the survival and development of the maintenance unit
A) 20,9% Positive due to its ability for international expansion
B) 79,1% Negative due to
a) 42,3 the governmental plan for privatisation
a) 36,9 the organisational separation with the Airline, its almost exclusive client
b) 20,8% both of the aforementioned causes
C) 3,5% Positive the Maintenance Unit development due to the abilities for international expansion, but negative on their job status due to worsen of their labor conditions.

1) The strong and long term opposition of the a/c engineers union against the privatisation

1) Their fear which is based on their international experience that:

a) their labor conditions will worsen due to privatisation.

b) their job security-stability will be lost due to the decrease of the Airline's network which will result at the decrease of their job positions which will not be replaced by a probable expansion of the Maintenance Unit works.

Conclusions

A) The considerations of the aircraft engineers of this particular Maintenance Unit regarding

a) the effectiveness and efficiency of the incentives applied by corporate top management

b) the incentives which they require to be applied by corporate top management, are affected by the following internal and external corporate environment factors:

1) The national culture and the corporate labor culture.
2) Their international experience
3) The unreliability of the state-shareholders' selections regarding the developmental prospects of the company
4) The awareness of the a/c engineers power (monopolistic position, 100% unionised)
5) The international official recognition of the superior quality of the results of the a/c engineers work on flight safety.
6) The union's consideration
7) The political ideology.

B) The need for a social recognition is very strong among the particular a/c engineers

C) The selection of the incentives which attracted the a/c engineers in their profession, those which they required to be applied by the corporate management and its prioritisaton we consider that are biased by

a) the a/c feelings arose by the failure of the ten years efforts of the government the Airline's exclusive shareholder, to make the Airline competitive

b) the organisational restructuring of the Airline's corporate group that took place during our research period of time. which the a/c engineers considered that it will have negative effect on their job status

Bibliography

Airline's C.E.O. (2004) Proceedings of the Airline's Aircraft Engineers Union congress "The Air Transport Policy for the next decade for the "Airline" and its maintenance base "April 29. Athens

European Union (1994) Recovery program of the Olympic Airways S.A. (October)

European Maintenance Management Academy (1999/10) Auditors approved by I.A.T.A.

Greek Parliament (2002) Statement of the minister of Transportation mr Verelis, in the 12th of April 2002 debate among the members of the "economic transparency committee" regarding the Olympic Airways

Herzberg, F., Mausner, B., Synderman, B.(1959) *The motivation to work* New York: John Wiley & sons Inc.

Herzberg, F. (1966). *Work and the Nature of Man.* Cleveland: World Publishing Co.

Herzberg, F.(2003) *One more time: How to motivate employees* Harvard Business Review on Motivating People p. 45-72

Hofstede, G (1980a Summer*). Motivation, Leadership and Organizations:Do American theories apply abroad?* Organizational Dynamics, p. 42-63

Hofstede, G (1980b). *Culture's consequences: International diff erences in work-related value*s.Beverly Hills , CA: Sage

Hofstede, G (1983 Fall). *The cultural relativity of organizational practices and theories.* Journal of International Business Studies,p. 75-89.

Hofstede, G (1991). *Cultures and Organizations: Software of the mind.* Maidenhead, England: McGraw-Hill

Johnston, A. N. (1993) *CRM: Cross-Cultural Perspectives* In E.L.Wiener,B.G.Kanki, R. L.

Lainos J. S. (1992) Is the privatisation the only way? File Olympic Airways Editions STAXY. Athens-Greece

Lainos J.S.&Nikolaidis E. (2003) *The flight crew motivation at a scheduled air carrier in a liberalise international institutional environment: The case study of the Greek flag carrier the Olympic Airways S.A.* Proceedings of the Aviation management Education and Research Conference. Montreal 2003

Lainos J.S. (2003) *Introduction of changes for the restructure of an international scheduled flag carrier in a liberalize international institutional environment The case study of* the Olympic Airways. Aviation management Education and Research Conference. Montreal-Canada

Lainos J.S. (2004**)** *Particularities of labor's productivity in an international scheduled airline. Case study: The change of the Olympic Airway's (Greece) average labor's productivity trend, compared with the I.A.T.A.'s 1975*-2002". Proceedings of the Annual ATRS congress Istanbul Turkey

Manos, S. (2001) Question to the Greek Parliament (8/9) regarding the overpricing by the Airline of the acquired 4 A-340 a/c

Maslow, A. (1954). *Motivation and Personality.* New York: Harper & Row.

Verelis, (2004) To Vima Greek daily gazette. Press conference of Mr Verelis the former minister of transportation (4/14)

EFFECTS OF HAPTIC FEEDBACK IN THE TELE-OPERATION OF AN UNMANNED AERIAL VEHICLE

T. Mung Lam, **Maxime Delannoy**, **Max Mulder**, **M. M. (René) van Paassen**

Delft University of Technology, Faculty of Aerospace Engineering, Control and Simulation Division
Delft, The Netherlands

This paper will describe an experiment that investigates the influence of force feedback on collision avoidance, control behavior and workload in the tele-operation of an unmanned aerial vehicle (UAV). Artificial force fields are used to provide force information. Subjects are asked to control a stability-augmented UAV helicopter through an obstacle-loaded environment. Visual information is provided by a display containing the simulated forward looking camera view and a navigation display, providing a top-down view. The force feedback algorithm is only implemented for the horizontal plane.

Problems related to the general principle of an artificial force field that occur with autonomous robots, such as difficult passage through closely-spaced obstacles or oscillatory motions of the vehicle might also occur here, and are represented by the stick motions. Various subtasks during the experiment are conducted to investigate whether these possible problems actually occur and how they affect the operator performance and workload.

The experiment results indicate that haptic feedback is very useful to assist the human tele-operator to avoid collisions, especially in cases where the visual information becomes insufficient. The minimum distance between the vehicle and an obstacle increases and the time spent within a critical distance towards an obstacle decreases, all leading to a higher level of safety.

Introduction

In the tele-operation of an unmanned aerial vehicle (UAV), the human operator is physically separated from the vehicle. This leads to the situation where the operator lacks the normally available, rich amount of information sources such as motion, tactile and auditory cues. Visual information provided by on-board cameras is dominantly used to provide the operator information about the environment. However, due to the limited camera field of view this visual information alone provides very limited situation awareness and may not be sufficient for a safe and efficient control of UAV (Diolaiti and Melchiorri, 2002, Hogan, Pollak and Falash, 2002). This can occur particularly in cases where the camera is not pointing in the direction of motion, such as in a hovering helicopter. Therefore, it is recommended to provide the operator with multi-sensory information.

Force feedback can be used to provide the operator tactile information that complements the visual information about the environment (Anderson and Spong, 1989, Elhaij, Xi, Fung, Liu, Li, Kaga and Fukuda, 2001). The integration of multi-sensory information allows an improvement of situation awareness. This paper describes an experiment, investigating the effect of haptic feedback on collision avoidance, control behavior and workload. It is structured as follows. First a brief review will be given of potential fields, mapping the environment constraints to virtual forces. Then the experiment will be described followed by the results and conclusion.

Potential Fields

In order to provide force feedback to avoid collision it is required for the control manipulator to provide the force before the vehicle actually makes contact with an obstacle (environment constraint). In literature potential fields are often used for local path planning of autonomous (ground) robots, mapping the environment constraints to the controller to avoid local obstacles (Borenstein and Koren, 1989, Khatib, 1986, Krogh, 1984). The obstacles exert virtual repulsive forces, pushing the robot away from the obstacles, whereas the goal at which the robot should arrive exerts an attractive force.

Two potential fields that are often referred in literature will be briefly discussed in this section, followed by a description of an artificial force field that was developed at this faculty.

Artificial Potentia Field

One of the first potential field was introduced by Khatib , which is called the Artificial potential field (Khatib, 1986). It depends only on the position of the system with respect to an obstacle and requires analytical description of obstacles. However, in unknown environment with complex obstacles, this field would not be suitable.

Generalized PotentialField

Krogh introduced a potential field that depends on the position as well as on the velocity towards an obstacle (Krogh, 1984), Figure 1. This type of field would be more representative for a level of danger. When the vehicle is close to an obstacle but moving away from the obstacle or parallel to an obstacle, the repulsive forces would not be large. On the other hand, when the vehicle approaches an obstacle with a large speed from a reasonable distance, the repulsive forces would be large. Additionally, the field also considers the acceleration limitation of the vehicle. However, from the application of potential fields for autonomous robots, some limitations were found. In case of closely space obstacles, the robot would move with oscillatory motions between these obstacles. A minimum in the potential field may occur, causing the robot to stop. Additionally, the generalized potential field may be too large for a reasonable velocity that may not be compatible with the operator's internal representation of the environment constraints.

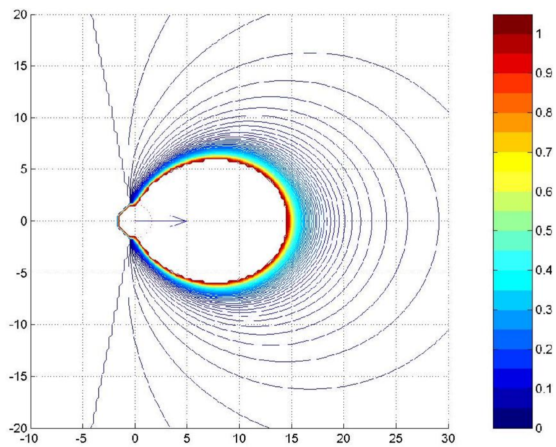


Figure 1. *Generalized potential field.*

Parametric Risk Field
Due to the challenges of the potential fields discussed above an artificial force field was developed, called Parametric risk field (Boschloo, 2004). The field is based on the principle of the generalized potential field, but it also allows the user to change its size and shape through certain parameter settings for certain tasks. Figure 2 hows a schematic presentation of the parametric risk field. Wit d_0 the width of the field can be adjusted, whereas with d_{ahead} the length of the field can be adjusted.

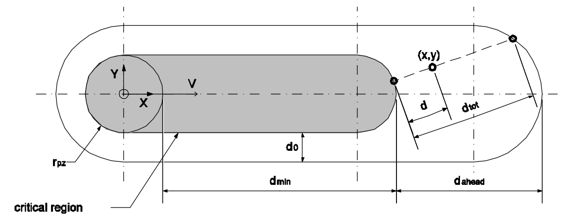


Figure 2. *Parametric risk field.*

A previous study (Boschloo, 2004) indicates that the parametric risk field can be used to avoid collision with less oscillatory motions with respect to the generalized potential field. An experiment conducted by Lam et al. (2004) showed that the parametric risk field can improve the path following performance considerably, at the cost of a higher workload.

However, the experiment involved a path following task through a tunnel-in-the-sky display of which the tunnel walls represent the environment constraints.
A more realistic experiment should be conducted having an UAV flying freely through an environment with obstacles.

Experiment

The goal of the experiment is to investigate the effect of haptic feedback on the collision avoidance, control activity and workload.

Subjects

Eight subjects with no flight experience participated in the experiment.
The main task for the subjects was to follow a trajectory without colliding with any obstacles. The trajectory contains different scenarios, in which a specific maneuver (subtask) needs to be conducted. Additionally, the scenarios are defined in such a way that they are similar to those that would introduce control difficulties for autonomous robots, found in literature.

Apparatus

The experiment was conducted in a fixed-base simulator in the Human-Machine Laboratory of the Control and Simulation division. A hydraulic driven side-stick was used to provide force feedback. Mass-spring-damper stick dynamics were simulated.

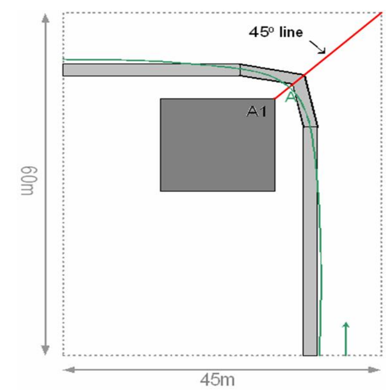
Independent Variables

There are three levels of haptic configurations (HC) and six levels of subtasks (ST). The haptic configurations are:

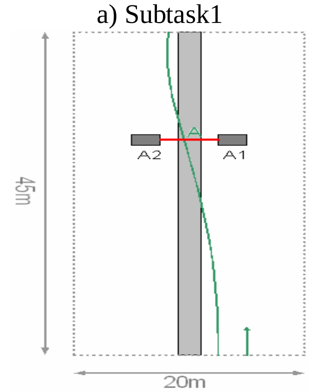
1. No haptic haptic feedback, i.e. the subjects only feel the simulated mass-spring-damper stick dynamics.
2. Basic risk field, i.e. force feedback generated by a slightly modified version of the generalized potential field.
3. Parametric risk field, i.e. force feedback generated by the recently developed force field. The parameter settings are d_0=1.5 m and d_{ahead}=2 x V, where V is the velocity.

Second, each of the subtasks will be described briefly below with the item number corresponding to the task number. The scenarios for each subtask are shown in Figures 3 and 4. In these pictures, the light gray path represents the reference trajectory, whereas the green line represents an example of a trajectory that the UAV might fly. The dark gray objects represent the obstacles.

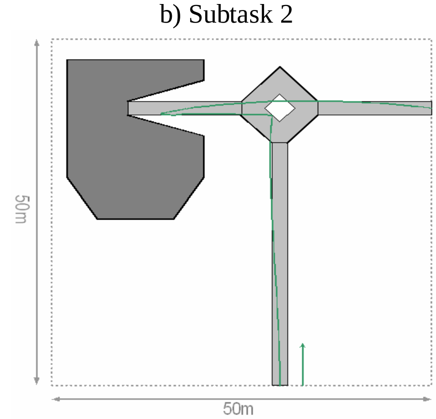
1. In this scenario the helicopter has to make a 90 degrees turn around a building. See Figure 3a. During the turn the building will be out of the camera field of view. It is expected that without haptic feedback corner-cutting effects may occur, leading to a larger amount of collisions than with haptic feedback. The length A-A1 is used to represent the minimum distance between the vehicle and obstacle.
2. In this scenario the helicopter is to fly between two closely-spaced, small obstacles. In literature, this scenario would lead to difficult or no passage with autonomous robots. It is expected that the operator needs more effort to push the helicopter through the passage, but less effort to avoid collision with either one of the obstacles. The smallest value between the lengths A-A1 and A-A2 represents the minimum distance to an obstacle. See Figure 3b.
3. This scenario demands a special task in a hovering phase of the helicopter. Once the helicopter has reached the square, it should hover backwards into the direction of A1 until the operator can see a certain stop sign fixed in the world. See Figure 3c. In this scenario, the camera visual information does not point in the direction of motion and it is expected that haptic feedback would become very useful in this kind of situations and tasks.
4. This scenario consists of a building with a discrete change in the shape of the wall. It is expected that this would lead to a discrete change in the force feedback, leading to a deviation from the reference path. See Figure 3d.
5. In this scenario two buildings with discrete changes in the opposite direction may lead to oscillatory behavior in the stick and cause considerable control difficulties. See Figure 4a.
6. In this scenario the turn radius with haptic feedback will be limited due to the obstacles in front and at the left side. It is expected that this scenario will lead to control difficulties, when approaching with high speed. See Figure 4b.



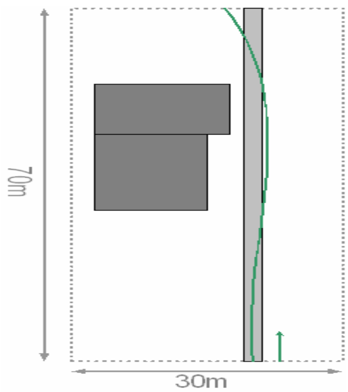
a) Subtask1



b) Subtask 2

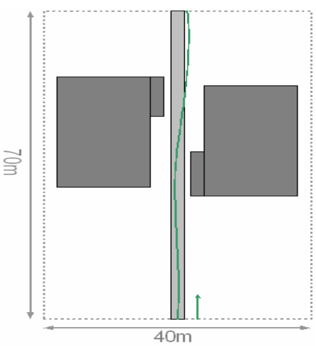


c) Subtask 3

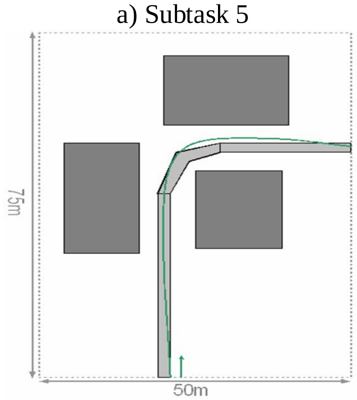


d) Subtask 4

Figure 3. *Subtasks 1 to 4.*



a) Subtask 5



b) Subtask 6

Figure 4. *Subtasks 5 and 6.*

Dependent measures

The efficiency of collision avoidance can be expressed by the number of collisions. The minimum distance with respect to an obstacle and the time spent within a critical distance to an obstacle are used as a measure for the level of safety. The standard deviation of the total exerted moment on the stick represents pilot control activity, whereas the standard deviation of the total external moment represents the haptic activity. The workload is measured by means of a TLX rating scale (Hart and Staveland, 1988).

Procedure

Each subject will fly 5 runs for each haptic configuration. Before the actual experiment, subjects get the opportunity to get familiar with the three haptic configurations by training runs. After each experiment run, subjects are asked to rate their workload using the NASA TLX rating scale.

Description of the Experiment Simulation

Display A simulated onboard camera outside visual, showing the world in a 3-dimensional fashion is projected on a large wall in front of the operator. The reference path is shown in the simulated world as a gray path on the ground, see Figure 5.

A 2-dimensional navigation display is presented on a 15 inch screen located in front of the operator between two operator seats, see Figure 6.

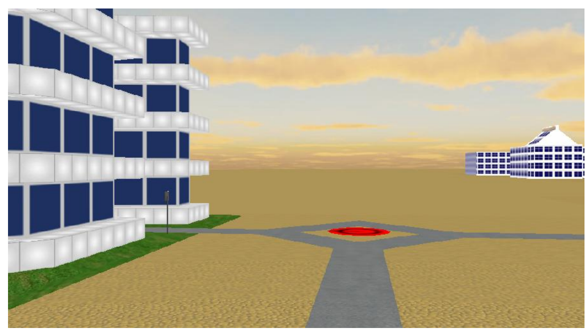


Figure 5. *Three-dimensional outside visual display.*

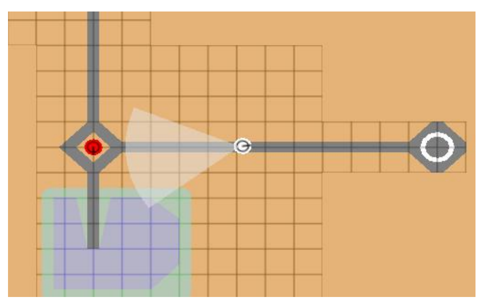


Figure 6. *Two-dimensional navigation display.*

Trajectory and helicopter model Five different trajectories are defined. Each trajectory contains three repetitions of the six scenarios in a random order. For each haptic configuration these five trajectories will be flown one time.

A stability-augmented UAV helicopter model with easy controllability is used. The model has a maximum velocity of 5 m/s and a maximum acceleration of 1 m/s^2.

Results and Discussion

The main results will be given in this section. A full-factorial ANOVA will be applied. The error bars, showing the mean and the 95% confidence intervals are shown in Figure 7.

Number of Collisions

A borderline significant effect of haptic configuration exists on the number of collisions, (HC: $F_{2,14}$=2.811, p=0.094). A Post-Hoc analysis (Student-Newman Keuls (SNK), α=0.05) reveals that in case of no feedback the most amount of collisions occur.

In subtasks 2 and 4 no collisions occur, which results in a significant effect of subtask, (ST: $F_{5,35}$=2.514, p=0.048) . Additionally, subtasks 3 and 5 lead to more collisions with no haptic feedback, resulting in a significant 2-way interaction (HC x ST: $F_{10,70}$=2.338, p=0.019).

Control Activity

A highly-significant effect of haptic feedback on the control activity was found. Independent of the subtask, the basic risk field causes the highest control activity, (HC: $F_{2,14}$=56.697, $p \leq 0.01$). A post-hoc analysis (SNK, α=0.05) revealed that the control activity is lowest with no haptic feedback and highest with the basic risk field.

Also a highly-significant effect of subtask was found, resulting in a high control activity in task 3 and low activity in tasks 2 and 4 with no haptic feedback and parametric risk field, (ST: $F_{5,35}$=16.966, $p \leq 0.01$). In subtask 3 the stick deflections are equivalent for the haptic configurations, in contrast with other subtasks. This expresses the highly-significant interaction, (HC x ST: $F_{10,70}$=22.564, $p \leq 0.01$).

Haptic Activity

The haptic configuration has a highly-significant effect on the haptic activity, (HC: $F_{2,14}$=211.024, $p \leq 0.01$). A post-hoc analysis (SNK α=0.05) showed that the basic risk field causes the highest haptic activity.

However, subtask 3 does not lead to a high haptic activity, causing a highly-significant effect of subtask, (ST: $F_{5,35}$=35.5, $p \leq 0.01$).

As can be seen in Figure 5b, subtasks 5 and 6 lead to higher haptic activity from the parametric risk field with respect to other subtasks, whereas it is not the case for the paremetric risk field. This causes a highly-significant interaction, (HC x ST: $F_{10,70}$=50.697, $p \leq 0.01$).

Minimum Distance from Obstacle

A highly-significant effect of the haptic configuration on the minimum distance from an obstacle exists, (HC: $F_{2,14}$=19.221, $p \leq 0.01$). A post-hoc analysis (SNK α=0.05) revealed that the basic risk field yields the largest distance, whereas no haptic feedback leads to the smallest distance to an obstacle.

In subtasks 2, 3 and 5 small distances occur with respect to other tasks. This expresses the highly-significant effect of subtasks, (ST: $F_{5,35}$=48.084, $p \leq 0.01$). For subtasks 2 and 3, the basic field does not yield the largest distance, expressing the highly-significant interaction, (HC x ST: $F_{10,70}$=17.488, $p \leq 0.01$).

Time Within Critical Distance

Only for subtasks 3, 5 and 6 the time can be measured, during which the helicopter is in a distance of 0.5 m or less from the obstacle.

A highly-significant effect of haptic feedback and subtask exist on the time, (HC: $F_{2,14}$=17.149, $p \leq 0.01$; ST: $F_{2.14}$=12.499, $p \leq 0.01$). For subtask 6 the difference between the haptic configuration, which expresses the significant interaction, (HC x ST: $F_{4,28}$ = 3.785, p=0.014).

Workload

Since the TLX is rated for a whole run, containing all subtasks, the workload cannot be distinguished for the different subtasks.

A highly-significant effect of haptic configuration leads to a highest workload by the basic risk field and the lowest workload in case of no haptic feedback, ($F_{2.14}$=39.717, $p \leq 0.01$).

From the six weightings, the physical demand, the effort and the frustration level play the greatest part in the high workload introduced by the haptic feedback.

Discussion

For simple subtasks such as in scenarios 2 and 4 no collisions occurred, independent of the haptic configuration. For scenarios 1 and 3, where the visual information becomes insufficient, the amount of collision can be reduced with haptic feedback. For complex subtasks in closely-spaced obstacles such as in scenarios 5 and 6 haptic feedback can even reduce the amount of collisions considerably.

Conclusions and Recommendations

Haptic feedback can assist the tele-operator to avoid collisions in complex tasks, where visual information becomes insufficient. Also the distance from the obstacle and the time spent within a critical distance are improved with haptic feedback, contributing to a higher level of safety. However, the reduction of collision and the improvement of the level of safety are at the cost of a higher workload and control activity.

Although it is shown that haptic feedback can improve the collision avoidance, it is unclear whether it can be related to an improvement of situation awareness. Therefore, it is recommended to employ a situation awareness assessment.

Information transportation time delay may well affect the collision avoidance performance and stability of the human-vehicle system. Time delay should be included in the system and investigated as well, in particular the effects on the biophysical feedback in narrow corridors.

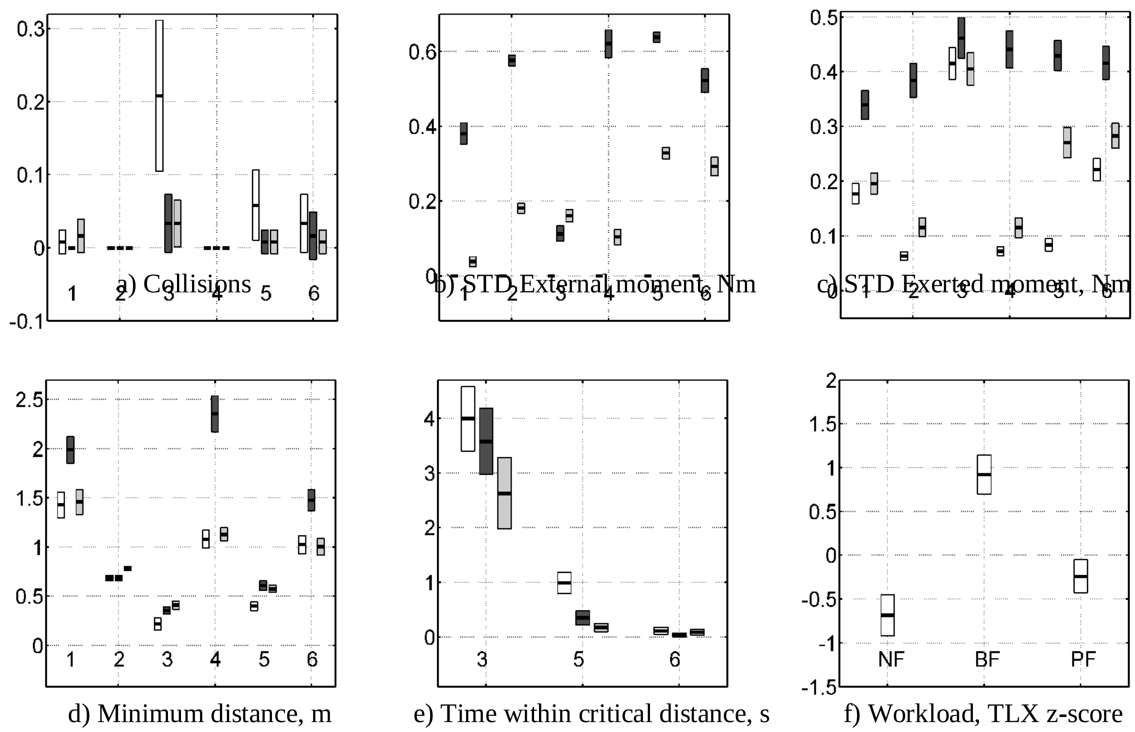


Figure 7. *The mean and 95% confidence limits. The numbers 1 to 6 represents the subtasks. The white, dark gray and light gray bars represent the no haptic feedback (NF), basic risk field (BF) and parametric risk field (PF), respectively. Note that in f) the error bars are categorized by haptic configuration.*

References

Anderson, R. J. and Spong, M. W. (1989). *Asymptotic stability for force reflecting teleoperators with time delay*, Proceedings of IEEE International Conference on Robotics and Automation, Vol. 3, pp. 1618-1625 Borenstein, J and Koren, Y. (1989). *Real-time obstacle avoidance for fast mobile robots*, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 19, No. 5, pp. 1179-1187

Boschloo, H. W., Lam, T. M., Mulder, M., van Paassen, M.M. (2004), *Collision Avoidance for a Remotely-Operated Helicopter Using Haptic Feedback*, Proceedings of IEEE International Conference on System, Man and Cybernetics, 2004, pp. 229-235

Delannoy, M (2004), *Experimental evaluation of two CAS systems for the remote control of UAV*, Unpublished Msc. Thesis report, Delft University of Technology

Diolaiti, N. and Melchiorri, C. (2002). *Tele-operation of a mobile robot through haptic feedback*, IEEE International Workshop 2002 HAVE, pp. 67-72

Elhaij, I., Xi, N., Fung, W. K., Liu, Y.H., Li, W. J., Kaga, T., and Fukuda, T. (2001). *Haptic information in internet-based teleoperation*, IEEE/ASME Transactions on Mechatronics, Vol. 6, No. 3, pp. 295-304

Hart, S. G. and Staveland, L. E. (1988). *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research*, in Human Mental Workload

Hogan, J., Pollak, E., and Falash, M. (2002). *Enhanced situational awareness for uav operations over hostile terrain*, AIAA's 1st Technical Conference and Workshop on Unmanned Aerospace Vehicles, No. AIAA 2002-3433

Khatib, O. (1986). *Real-time obstacle avoidance for manipulators and mobile robots*, The International Journal of Robotics Research, Vol. 5, No 1., pp. 90-98

Krogh, B. H. (1984). *A generalized potential field approach to obstacle avoidance control*, Society of manufacturing Engineers Technical Paper MS84-484

Lam, T. M., Boschloo, H. W., Mulder, M., van Paassen, M.M., van der Helm, F.C.T. (2004), *Effect of Haptic Feedback in a Trajectory Following Task with an Unmanned Aerial Vehicle*, Proceedings of IEEE International Conference on System, Man and Cybernetics, 2004, pp. 2500-2506

A NON-LINEAR RELATIONSHIP BETWEEN CONTROLLER WORKLOAD, TASK LOAD, AND TRAFFIC DENSITY: THE STRAW THAT BROKE THE CAMEL'S BACK

Paul U. Lee, Joey Mercer
SJSU / NASA Ames Research Center
Moffett Field, CA

Nancy Smith, Everett Palmer
NASA Ames Research Center
Moffett Field, CA

Controller workload, recognized as a significant bottleneck to capacity increase in the future National Airspace System, has been researched extensively in air traffic management. Unfortunately, subjective workload has been an unreliable predictor of a controller's ability to safely manage the traffic, leading to attempts at replacing workload with more objective metrics, such as task load (e.g. number of clearances) and traffic density (e.g. aircraft count). A significant caveat to substituting these metrics for workload ratings, however, is that their relationships are non-linear. More specifically, when the objective metrics, such as aircraft count, increase linearly, the controller's perceived workload remains low until the traffic and associated task load increase to a critical threshold. From this point, the workload increases at a much faster rate with each added task. In a series of informal studies conducted as a precursor to testing Distributed Air Ground Traffic Management (DAG-TM) concepts, researchers at NASA Ames Research Center manipulated aircraft count in real-time human-in-the-loop simulations to determine the maximum traffic levels at which the controllers stated that traffic would no longer be manageable. As hypothesized, traffic scenarios that elicited moderate levels of controller workload quickly became unmanageable when only a few aircraft were added. Feedback from the controllers further supported the non-linear nature of subjective workload. Task load data partially supported the above findings but the results were inconclusive due to individual differences and varying results from different task load metrics. The non-linear relationship between subjective workload and aircraft count has been further examined using data collected from the Free Maneuvering concept feasibility study in June 2004, which shows a step-function relationship between the two. The combined results suggest that any estimation on workload should not be extrapolated linearly from a set of workload measures taken from an experiment since the extrapolated workload is likely to significantly underestimate workload.

Introduction

Controller workload has been a focal topic in air traffic management research (e.g. Stein 1985, Athenes, Averty, Puechmorel, Delahaye, and Collet, 2002). It is considered to be a key limiting factor to capacity increase in future air traffic operations. However, subjective workload has many undesirable characteristics. First, workload ratings have shown to have significant individual differences, making them difficult to be used as a reliable metric that can be generalized to different sectors and controllers. Furthermore, while objective metrics can be derived from traffic and sector characteristics, workload ratings are derived only after controllers work the traffic, making them difficult to use as a predictive metric that can prevent future traffic overload.

One potential solution to this problem is to replace subjective workload metrics with correlated objective metrics, such as peak aircraft count, traffic geometry, total time in sector, number of clearances, etc. A general approach to correlating workload with objective metrics is to identify and/or define factors that are likely to correlate with workload, use multivariate linear regression models to fit the data, and then eliminate factors that contribute little to workload prediction. From these types of analyses, peak aircraft count has generally emerged as one of the best predictors of workload (e.g. Manning, Mills, Fox, Pfleiderer, Mogilka, 2001).

Most of these analyses assume linear correlation between workload ratings and objective metrics. This assumption seems to run counter to the subjective experience of workload. Controllers often report a low to moderate level of workload for a seemingly busy traffic but at some point report much higher workload with few added tasks and/or minor off-nominal events. In general, there seems to be a non-linear relationship between workload and objective metrics. A controller may *perceive* the workload to be low until the traffic and associated task load reach a critical point, after which s/he perceives the workload to be high.

We examined the non-linearity of workload using data that was collected during an informal "traffic load test" which established the maximum traffic that a controller can handle with advanced decision support tools. Despite the informal nature of the study, the data provide some evidence and insight into the relationship between workload, aircraft count, and other task loads.

Method

Participants

Two certified professional air traffic controllers and two retired controllers/ supervisors participated in the study.

Tool Capabilities

Advanced air and ground-side decision support tools (DSTs) were integrated with Controller Pilot Data Link Communication (CPDLC) and the Flight Management System (FMS). This integration allows the controllers and the pilots to exchange 4-D trajectory information quickly and with low workload. The controller DSTs have been integrated into a high fidelity emulation of the Display System Replacement (DSR) controller workstation. In order to support the then tested concept, all aircraft were equipped with CPDLC, FMS, and automatic dependent surveillance-broadcast (ADS-B).

Airspace

The simulation airspace included portions of Albuquerque Center (ZAB), Kansas City Center (ZKC), Fort Worth Center (ZFW), and Dallas-Fort Worth TRACON (Figure 1). Arrivals transitioned Amarillo high and Wichita Falls high from the northwest and Ardmore high from the north. The two main streams of arrivals merged at the BAMBE meter fix in the Bowie low sector. The traffic mix in Amarillo consisted of arrivals and overflights in level flight. A significant portion of Wichita Falls traffic was arrivals while Ardmore had arrivals, departures, as well as a significant number of overflights.

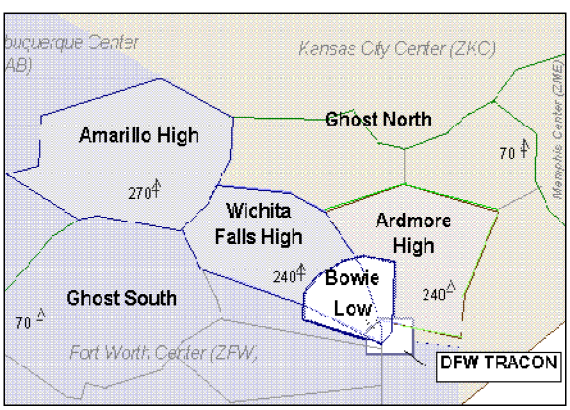


Figure 1. *Simulated airspace*

Procedure

The "traffic load test" was conducted to determine the maximum traffic levels that a controller can handle in each of the high altitude sectors. Each simulation run consisted of 30 – 40 minute traffic scenarios, in which the traffic gradually increased during the first fifteen minutes to a peak aircraft count and which was then sustained for the rest of the run. Ten versions of traffic scenarios were generated per sector, at an increment of two aircraft during the peak traffic.

Each sector – Amarillo, Ardmore, and Wichita Falls – was tested one at a time. Each controller participant was paired up with a supervisor who doubled as a support controller who handled the surrounding traffic that entered or exited the test sectors. The controller participants simultaneously worked the same sector in separate parallel simulated airspaces.

The controllers were given a briefing about the purpose of the study and were given training to familiarize themselves with the tools, traffic scenarios, and the overall procedures. After two days of training, the participants and the researchers discussed the definition of "unmanageable" traffic to arrive at a consensus on a common definition prior to starting the data collection runs.

For the data collection runs, a traffic level was picked based on the amount of traffic that was effectively handled during the training sessions. After working the traffic at the initial traffic level, the controller participants and the supervisors discussed and came to a group consensus on the traffic level with respect to their ability to effectively control the traffic. If they thought that the traffic was below the maximum traffic level, they worked another traffic scenario that increased the traffic by four peak aircraft count, and then evaluated the traffic after the run. If they thought the new traffic level was unmanageable, the peak traffic count was decreased by two. The decision process repeated until the maximum traffic level was established. If the traffic was impossible to work at any time, they had the option to stop the simulation run at any time. This procedure was modeled after the staircase method of establishing thresholds in psychophysical measurements. (Cornsweet, 1962)

Results & Discussion

Definition of "Unmanageable" Traffic

After the training and prior to data collection, participants were asked what they would consider as "unmanageable" traffic. Surprisingly, there was a remarkable agreement among the participants in their assessments. They generally agreed that the traffic is unmanageable once they lose their situational awareness of the aircraft. They also described this as losing the "flick". They described having the "flick" as having the "picture", a plan to work the traffic proactively to provide traffic management rather than reactively to avoid conflicts. They felt that once they lost the "flick", they have already compromised safety even if it did not result in any operational errors.

Some of the potential indicators that a controller is near the maximum traffic level are:

- handoffs are late
- can't find check-in flights easily
- reactive instead of proactive traffic control
- don't know where the planes are
- situation startles you
- service goes out the window

One controller remarked that when the traffic reached unmanageable levels during training, he was startled to "see" an aircraft for the first time in the middle of his sector heading for another plane. Luckily, the planes were separated by altitude but it would have resulted in a separation loss otherwise.

They also commented that near the maximum traffic level, a controller might feel that s/he is fine but one more problem – even something as simple as an altitude request – may put him/her "down the tubes". Supervisors commented that part of their job is to recognize when a controller might have reached his/her workload threshold so that they can provide relief or help before the person goes "down the tubes." They utilize the controller's body language, speech, etc., as cues for help.

Aircraft Count

The controllers worked various traffic levels during training, which allowed them to quickly converge on the maximum traffic level during data collection. As hypothesized, a small change in the aircraft count had a significant impact on the controller workload when the traffic was near the maximum.

For Ardmore and Wichita Falls sectors, three levels of workload – moderate, maximum, and unmanageable – were reported during data collection. As shown in Figures 2 and 3, the number of aircraft that was controlled was very similar between the scenarios reported as moderate and maximum levels of workload. The peak aircraft count was slightly higher in scenarios that the participants reported as unmanageable workload.

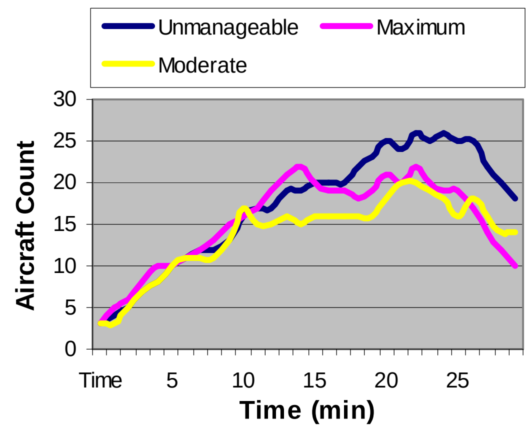


Figure 2. *Controller-owned aircraft in Ardmore*

The difference in aircraft count from moderate to unmanageable workload was relatively low – i.e. between 4 to 5 aircraft – suggesting that workload measurements were sensitive to minor changes in aircraft count. For the Ardmore sector, the average aircraft count during the ten minute peak was 17.2, 19.9, and 22.7 aircraft for moderate, maximum, and unmanageable workload, respectively. For the Wichita Falls sector, the average was 15, 14.7, and 18.7 for moderate, maximum, and unmanageable workload, respectively.

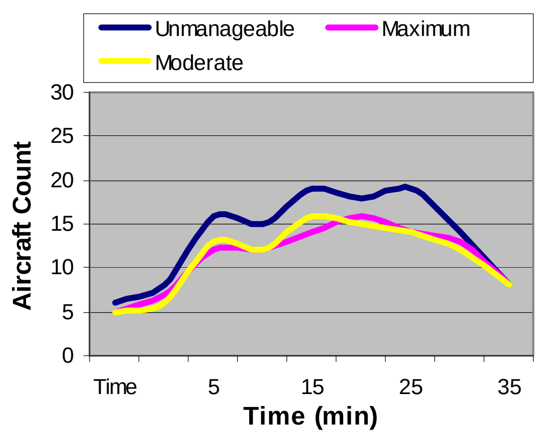


Figure 3. *Controller-owned aircraft in Wichita Falls*[1]

[1] Due to data logging problems, aircraft count was logged at every five minutes for this sector.

It is unclear why the moderate and maximum traffic levels had similar aircraft count in Wichita Falls sector. The task load data showed that controllers accepted more handoffs (four) and issued more clearances (3 – 11) in the maximum traffic scenario, suggesting that there were some measurable differences between the two scenarios. Further analysis is needed to understand the discrepancies between task load and aircraft count in this sector.

Figure 4 shows the number of aircraft controlled in unmanageable and maximum traffic scenarios in Amarillo sector. The maximum and unmanageable traffic had 21 and 23 aircraft during the peak ten-minute duration, respectively.

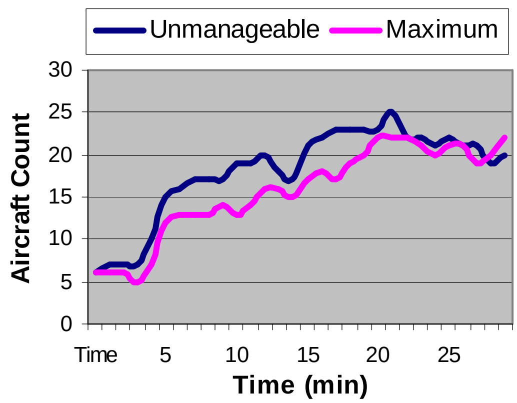


Figure 4. *Controller-owned aircraft in Amarillo*

Unfortunately, moderate traffic scenarios were run during the training sessions but not during data collection in this sector. Similar to Ardmore and Wichita Falls sectors, controllers reported a relatively moderate workload for traffic scenarios slightly below the threshold traffic, suggesting that workload increases from moderate to unmanageable with few additional aircraft.

Although the data from the load test suggest a large change in perceived workload with a small change in aircraft count, they do not directly demonstrate non-linearity in workload. However, a subsequent DAG-TM study provided more direct evidence of non-linearity. Figure 5 illustrates the non-linear relationship between workload and aircraft count. During the DAG-TM study, controller participants reported workload every five minutes during the simulation runs using a Workload Assessment Keyboard (WAK) on a scale of 1 to 7 (Stein 1985). For the four simulation runs that contained maximum controller-managed traffic levels, these ratings were correlated with peak aircraft count during the corresponding five minute duration. The observed data in Figure 5 shows an example of the non-linearity in Amarillo sector. Reported workload was low for an aircraft count up to 16 and then quickly ramped up to high workload from 16 to 22 aircraft. An S-curve, estimating a step function in workload from low to high, provided a better fit to the observed data than a linear or an exponential regression line/curve (a complete analysis is in Lee, submitted). The results suggest that subjective workload is categorical.

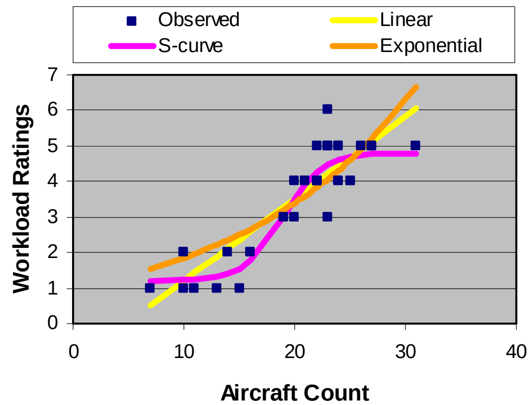


Figure 5. *Workload vs. aircraft count: observed and regression fits for Amarillo High*

Task Load

Controller workload was also compared to various task load metrics. A non-linear relationship between workload and task load metrics implies that small changes in task load would have resulted in large changes in workload. While some of the data supported this hypothesis, others were inconclusive.

Task load metrics were divided into three main categories: handoffs, clearances, and monitoring tasks, reasons of which will be described later. The analyses also kept the two controller participants' performance separate due to some interesting individual differences. Although task load analyses were done for all three sectors, we will focus mainly on Ardmore results due to space limitations, and selectively bring in results from the other two sectors. Overall, the pattern of results was similar for Ardmore, Amarillo, and Wichita Falls.

The number of handoffs that a controller accepts from an upstream sector and initiates to a downstream sector is directly related to number of aircraft in their sector. Figure 6 shows that for Ardmore sector, both controllers handled a near identical number of aircraft, and the number of handoffs initialized/accepted was, on average, 58, 72, and 80 for moderate, maximum, and unmanageable workload, respectively. For Wichita Falls, they were 61, 73, and 77 and for Amarillo, they were 69 and 73 for maximum and unmanageable workload. In all three sectors, the increase in the number of accepted handoffs between each traffic level were quite small (2 – 5), confirming that number of aircraft that the controllers worked were quite similar between moderate, maximum, and unmanageable traffic scenarios.

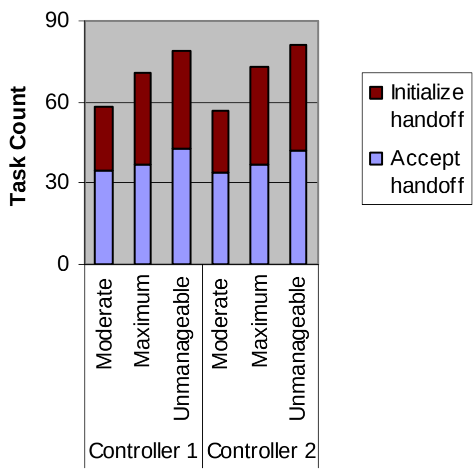


Figure 6. *Number of handoffs initiated and accepted for Ardmore sector*

The number of clearances that a controller issues may be a better indicator of controller workload since it addresses not only the traffic volume but also the traffic complexity. If an aircraft flying through a sector does not increase sector traffic complexity, controller may not need to issue any clearances to the aircraft. Figure 7 shows a count of speed and route clearances that were data linked to the flight deck, as well as altitude clearances issued by voice. There were additional speed and vector clearances by voice that were not analyzed and therefore excluded in this analysis. However, over-the-shoulder observation confirmed that there were very few voice-issued vectors or speed clearances due to easy uplink of speed and 4-D route clearances via data link using advanced DSTs.

Although aircraft count data indicated a similar number of controller-owned aircraft in moderate and maximum traffic scenarios (see Figure 2), the number of clearances were greater in maximum (32 for controller 1; 40 for controller 2) than in moderate traffic (22 for controller 1; 32 for controller 2). Therefore a large increase in controller workload between moderate and maximum scenarios may be better explained by the number of clearances than the aircraft count. However, a lack of distinct difference between the number of clearances in the maximum and unmanageable traffic scenarios limits its ability to fully explain its relationship to workload. In addition, the clearance data from Wichita Falls and Amarillo sectors indicate only a modest increase (1 – 5) in the number of clearances between traffic levels in all but one instance. There were also individual differences between the two controllers, as controller 1 issued fewer clearances than controller 2.

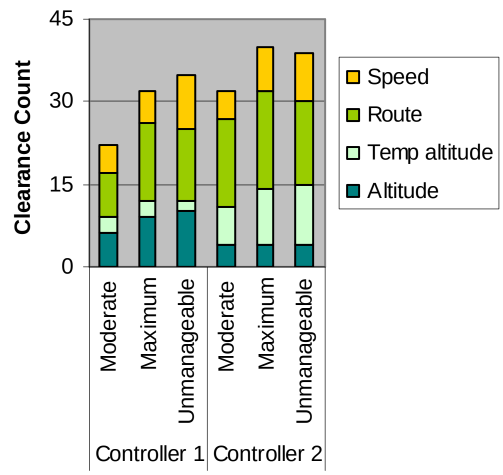


Figure 7. *Number of speed, altitude, and route clearances for Ardmore sector*

Controllers also engaged in various monitoring tasks. Most of the monitoring tasks were not recorded by the data collection system, but the ones that were logged show an interesting individual difference between the two controllers. Figure 8 shows the number of times the controller participants toggled or adjusted the data tags, displayed FMS routes, and displayed J-ring around the targets.

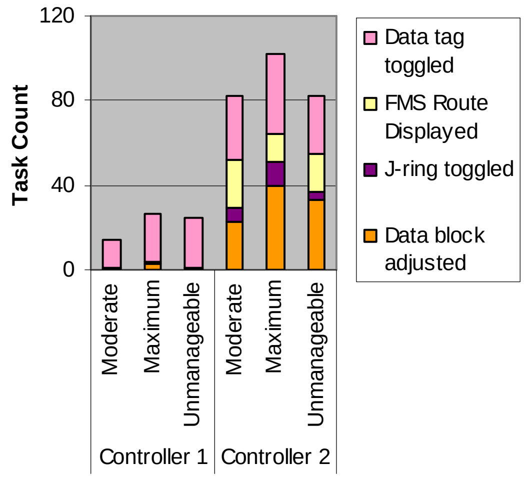


Figure 8. *Number of tasks associated with monitoring for Ardmore sector*

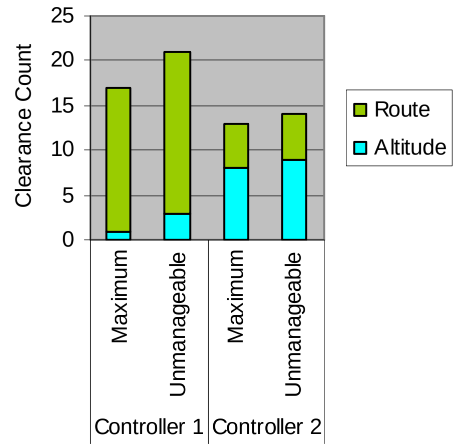


Figure 9. *Number of altitude and route clearances for Amarillo sector*

Data tag toggles and adjustments were often used as memory aids to let the controllers visually discriminate between aircraft that have been handed off, need to be attended to, etc. Display of FMS routes allowed them to verify where the planes were going, especially since the airspace and the traffic scenarios were unfamiliar to them. J-rings were often used as additional memory aides as well as to visually emphasize the 5 nm separation boundaries for aircraft that had potential conflicts with other nearby aircraft.

As shown in Figure 8, there was a large difference in these types of activities between the two controllers in Ardmore sector. Similarly in Amarillo and Wichita Falls, controller 2 consistently engaged in more monitoring activities than controller 1. Controller 2 also engaged in less monitoring activities in unmanageable than in maximum traffic scenarios across all three sectors, perhaps because monitoring activities were lower priority tasks that were dropped when the controller became too busy. Overall, it is interesting that these types of activities did not seem to affect their overall workload assessment since the two controller participants generally agreed on their workload in each traffic scenario despite having a large difference in these monitoring activities.

Finally, one interesting finding unique to Amarillo sector was an individual difference in the types of clearance issued by the two controller participants. As shown in Figure 9, controller 1 issued mostly lateral route amendments while controller 2 issued more altitude clearances.

Controllers have commented that they try to resolve the conflicts using lateral maneuvers because 1) aircraft may be flying its preferred altitude and 2) an altitude maneuver is reserved as an "out" maneuver in case lateral maneuvers do not resolve the conflict. The data suggest that different controllers use different amount of lateral vs. vertical maneuvers in similar traffic situations.

Conclusion

There are interesting implications to the non-linear relationship between subjective workload and traffic count. First, any estimation on workload should not be extrapolated linearly from a set of workload measures taken from an experiment since the extrapolated workload is likely to significantly underestimate workload. The potential for underestimation of workload is greatest when evaluating future air traffic concepts that rely on automation to reduce task loads and increase capacity. Secondly, metrics such as traffic count or task loads should not be used interchangeably with subjective workload unless a better characterization of their relationship is established. Finally, non-linearity of workload implies the importance of determining the critical traffic levels that shift perceived workload from one level or category to another. This will be a significant challenge due to individual differences in controllers' abilities and off-nominal events that can critically affect the workload. Further research is needed to understand how to accurately account for these factors.

References

Athenes, S., Averty, P., Puechmorel, S., Delahaye, D., and Collet, C. (2002). ATC Complexity and Controller Workload: Trying to Bridge the Gap, *HCI-Aero 2002: International Conference on Human-Computer Interaction in Aeronautics*, Cambridge, MA.

Cornsweet, T. N. (1962). The staircase method in psychophysics. *American Journal of Psychology*, *75*, 485-491.

Lee, P. U. (submitted). A non-linear relationship between controller workload and traffic count, *49th Annual Meeting of the Human Factors and Ergonomics Society,* Orlando, FL.

Manning, C., Mills, S., Fox, C., Pfleiderer, E., and Mogilka, H. (2001). The Relationship Between Air Traffic Control Communication Events and Measures of Controller Taskload and Workload. *The Fourth International ATM R&D Seminar ATM-2001,* Santa Fe, NM.

Stein, E. (1985). Air Traffic Controller Workload: An Examination of Workload Probe, *DOT/FAA/CT-TN84/24.* DOT/FAA Technical Center. Atlantic City, NJ.

THE IDENTIFICATION OF TRAINING NEEDS FOR DEVELOPING AERONAUTICAL DECISION MAKING TRAINING PROGRAMS FOR MILITARY PILOTS

Wen-Chin Li; **Don Harris**
Cranfield University
Bedfordshire, United Kingdom

Chung-San Yu
Air Force Academy
Kaohsiung, R.O.C.

This research applies Human Factors Analysis and Classification System (HFACS; Wiegmann & Shappell, 2003) analyzing aviation accidents in the R.O.C. Air Force between 1978 and 2002 in order to identify the training needs of aeronautical decision-making (ADM). There were 523 accidents associated with 1762 human errors. The results indicated that decision errors had been involved in 223 (42.6%) accidents. Without in-depth analysis of decision errors in military aviation, it is unlikely to identify precisely the training needs of ADM and the nature of the training content required to prevent the decision errors in aviation (Patrick, 2003). This research found that 'decision-errors' has significant association with lieutenant pilots and at landing phase, and pilots at the rank of 'cadet' (experience) flying 'training aircraft' (tools) practicing 'close pattern' (missions) at 'landing phase' (working environment) with the highest probability of accidents. It is important to understanding the junior pilots were very vulnerable to the decisions and supervisions made by high-level management. As Dekker (2001) described that human errors is systemically connected to the tools, tasks, and operational and organizational environment of operators, it is important to clarify the role of decision errors in pilot's tools, tasks, experience, and operating environment in military aviation in order to develop effective ADM training programs for military pilots.

Introduction

Identification of ADM Training Needs

Decision making performance in the aviation domain is a joint function of the features of the tasks and the pilots' knowledge and experience relevant to those tasks (Orasanu & Connolly, 1993). Orasanu (1993) has pointed out that no evidence exists to support the development of training techniques to improve all-purpose decision making skills. There are six different component skills involved in the six different types of decisions.

For improving aviation safety, it is important to identify the training needs for ADM. The Interservices Procedures for Instructional System Development (IPISD, Branson et al., 1975) was developed in the context of US military training. The intention was to disseminate principles concerning the development of training programs. The IPISD model divided the development of training into five main phases: analyze, design, develop, implement and control. Without accurate analysis, it is not possible to identify the ADM training needs and the content of training programs required for preventing aviation accidents. There are two general types of analysis techniques: task analysis which is used in training development for analyzing the knowledge, skills, and attitudes required by the operator in order to execute the task efficiently, and error analysis which focuses on errors in task performance (Patrick, 2003). Both analysis techniques can help to identify training needs and training content. The first type of task analysis is described as a traditional form of job/task/cognitive analysis. It breaks down work into a series of subtasks that have to be accomplished in a logical fashion. The second type, error analysis, is used to identify where training can be profitably directed for curing weaknesses. Dekker (2001) has proposed that human errors are systematically connected to features of the operators' tools and tasks, and error has its roots in the surrounding system. Analyzing incidents or accident reports can obtain a great deal of valuable information for identifying training needs for subsequent training to mitigate human errors.

Human Factors Analysis and Classification System

HFACS is a generic human error framework originally developed for US military aviation as a tool for the investigation and analysis of the human factors aspects of accidents. HFACS is based on Reason's (1990) system-wide model of human error in which active failures are associated with the performance of front-line operators in complex systems and latent failures are characterized as inadequacies or mis-specifications which might lie dormant within a system for a long time and are only triggered when combined with other factors to breach the system's defenses. These latent failures are spawned in the upper management levels of the organization and may be related to manufacturing, regulation and/or other aspects of management. As Reason (1997) noted, complex systems are designed, operated, maintained, and managed by human beings, so it is no surprising that human decisions and actions are implicated in all organizational accidents.

HFACS examines human error in flight operations at four levels. Each higher level affects the next downward level in HFACS framework.

- Level-1 'Unsafe acts of operators': This level is where the majority of causes of accidents are focused. Such causes can be classified into the two basic categories of errors and violation. Decision-errors are in this level.
- Level-2 'Preconditions for unsafe acts': This level addresses the latent failures within the causal sequence of events as well as more obvious active failures. It also describes the context of substandard conditions of operators and the substandard practices they adopt.
- Level-3 'Unsafe supervision': This level traces the causal chain of events producing unsafe acts up to the front-line supervisors.
- Level-4 'Organizational influences': This level encompasses the most elusive of these latent failures, fallible decisions of upper levels of management which directly affect supervisory practices, as well as the conditions and actions of front-line operators (Shappell & Weigmann 2001; 2003 & 2004; and Weigmann & Shappell 1997; 2001a; 2001b; 2001c & 2003).

Between 1996 and 2000, the Republic of China (R.O.C.) Air Force converted from the F-104 to a series of new generation fighters including F-16, Mirage 2000-5 and the self-developed IDF. To improve flight safety, R.O.C. Air Force Headquarters investigate the pattern of mishaps annually. The findings are that accidents attributable solely to mechanical failure decreased markedly in the recent years, but the contribution of human error has declined at a slower rate. Jensen and Benel (1977) found that decision errors contributed to 35% of all nonfatal and 51% of all fatal general aviation accidents in the United States between 1970 and 1974. Diehl (1991) following Jensen and Benel's research found that decision errors contributed to 56% of accidents in airlines and 53% of accidents in military aviation between 1987 and 1989.

In order to improve aviation safety there is a need for military pilots to be trained in making decisions related directly to the specific tactical environment. However, there is no research on the identification of training needs for aeronautical decision-making (ADM) and for developing the content of training programs for military pilots in the R.O.C. Air Force. This study applies the Human Factors Analysis and Classification System (HFACS) for analyzing human factors accident data from the R.O.C. Air Force. For developing effective ADM training programs, it is necessary to understand the association of decision errors with pilots' tools (aircraft), tasks (missions), ranks (flying experience), and flight stages (environment).

Method

Data

The data were comprised of the narrative descriptions of accidents occurring in the R.O.C Air Force between 1978 and 2002. In total, the complete data set comprised 523 accidents in this 25 year period.

Demographic Variables

This investigation analysed each accident using the following demographical variables.

1. Type of aircraft: the types of aircraft involved in accidents included fighters (F16, M2000, IDF, F104, F-5, etc.), cargo aircraft (B1900, C130, C123, C47, etc.), and training aircraft (AT3, T34, etc.).
2. Missions: accidents occurred when pilots' were performing missions that included air interception, air combat tactics, instrument flight, cross country, transition, surface attack, close pattern, test flight, and exercise.
3. The flight stages in which accidents occurred included: taxi before take-off, take-off, climb-out, flight in the operational area, decent, approach, landing and taxi after landing.
4. The ranks of pilots involved in accidents included: cadet, lieutenant, first lieutenant, captain, major, and lieut. colonel (above).

Classification Framework

This study used the HFACS framework as described in Wiegmann & Shappell (2003). The first level of HFACS categorizes events under the general heading of 'unsafe acts of operators' that can lead to an accident including and comprises of four sub-categories of 'decision errors'; 'skill-based errors'; 'perceptual errors' and 'violations'. The second level of HFACS concerns 'preconditions of unsafe acts' which has a further seven sub-categories of 'adverse mental states'; 'adverse physiological states'; 'physical/mental limitations'; 'crew resource management'; 'personal readiness'; 'physical environment', and 'technological environment'. The third level of HFACS is 'unsafe supervision', including 'inadequate supervision'; 'planned inappropriate operation'; 'failure to correct a known problem', and 'supervisory violation'. The fourth and highest level of HFACS is 'organizational influences' and comprises of the sub-categories of 'resource management'; 'organizational climate' and 'organizational process'.

Coding Process

Each accident report was coded by two investigators, an instructor pilot and an aviation psychologist. These two investigators were trained on the HFACS framework together for 10 hours to ensure that they achieved a detailed and accurate understanding to the categories of the HFACS. They then analyzed each accident report independently. To avoid over-representation from any single accident, each HFACS category was counted a maximum of only once per accident. The count acted simply as an indicator of presence or absence of each of the 18 categories in a given accident.

Results

Sample Characteristics

A total of 523 accidents were analyzed. In these accidents, 1,762 instances of human error were recorded within the HFACS framework. Initial results found that acts at the level of 'unsafe acts of operators' were involved in 725 (41.1%) of instances; the 'preconditions for unsafe acts' level was as a causal factor in 552 (31.3%) of cases; the 'unsafe supervision' level was involved in 221 (12.5%) instances, and the 'organizational influences' level in the model was involved as a factor in 264 (15 %) cases. Decision errors were involved in 223 (42.6%) accidents. The inter-rater reliabilities assessed using Cohen's Kappa varied between 0.440 and 0.826, a range of values spanning between moderate and substantial agreement. Fourteen HFACS categories exceeded a Kappa of 0.60, which indicates substantial agreement. Four categories had Kappa values of between 0.40 and 0.59 indicating moderate levels of agreement (Landis & Koch, 1977) (table 1).

Effect of Aircraft Type

At the level of 'unsafe acts of operators', there were no significant associations with aircraft type. At the level of 'preconditions for unsafe acts', the associations of aircraft type with 'adverse mental states', 'crew resource management', and 'personal readiness' were significant. Training aircraft were over-represented in having 'adverse mental states' and 'personal readiness'; cargo aircraft were over-represented in having 'crew resource management' problems, even though the frequency of fighters was the highest. At the level of 'unsafe supervision', the associations of aircraft types with 'inadequate supervision' and 'failed to correct a known problem' were significant. Training aircraft were over-represented in these two categories of accidents. At the level of 'organizational influences', the association of aircraft types with 'organizational process' was significant. Training aircraft were over-represented in the category of 'organizational process' of accidents (see table 2).

Table 1. *The frequency and percentage of accident and reliability of HFACS categories*

HFACS Categories	Accidents' Frequency Percentage, and reliability		
	Frequency	Percentage	Cohen's Kappa
Organizational process	76	14.5%	0.593
Organizational climate	4	0.8%	0.440
Resource management	184	35.2%	0.768
Supervisory violation	8	1.5%	0.694
Failed correct known problem	12	2.3%	0.548
Planed inadequate operations	24	4.6%	0.706
Inadequate supervision	177	33.8%	0.826
Technology environment	44	8.4%	0.608
Physical environment	74	14.1%	0.797
Personal readiness	29	5.5%	0.695
Crew resource management	146	27.9%	0.801
Physical/mental limitation	73	14.0%	0.691
Adverse physiological states	2	0.4%	0.441
Adverse mental states	184	35.2%	0.748
Violations	160	30.6%	0.695
Perceptual errors	116	22.2%	0.667
Skilled-based errors	226	43.2%	0.712
Decision errors	223	42.6%	0.675

Effect of Aircraft Mission

At the level of 'unsafe acts of operators', the association of mission with 'skill-based errors' was significant. The 'close pattern' mission was over-represented in the category of 'skill-based errors' of accidents. At the level of 'precondition for unsafe acts', the association of mission with 'personal readiness' was significant. The 'close pattern' mission was also over-represented in the category of ‘personal readiness’ of accidents. At the level of ‘unsafe supervision’, the association of mission with ‘inadequate supervision’ was significant. Again, the ‘close pattern’ mission was over-represented in the category of ‘inadequate supervision’ of accidents. However, at the level of ‘organizational influences’, there was no significant association between mission and categories in the HFACS framework (see table 2).

Table 2. *The significant association between HFACS categories and demographical variables*

HFACS Categories	Significant association with HFACS categories			
	Types of aircraft	Missions Of pilots	Stages of flight	Ranks of pilots
Organizational process	χ^2=7.74, df=2, p<0.02			χ^2=11.1, df=5, p<0.05
Organizational climate				
Resource management				
Supervisory violation				
Fail correct problem	χ^2=20.6, df=2, p<0.00			
Plan inadequate operation				
Inadequate supervision	χ^2=8.28, df=2, p<0.01	χ^2=20.2, df=8, p<0.01	χ^2=34.6, df=8, p<0.00	χ^2=26.6, df=5, p<0.00
Technology environment				
Physical environment				χ^2=15.1, df=5, p<0.01
Personal readiness	χ^2=9.58, df=2, p<0.01	χ^2=23.1, df=8, p<0.01		
CRM	χ^2=8.35, df=2, p<0.01		χ^2=19.6, df=8, p<0.01	
Phy./mental limitation			χ^2=17.5, df=8, p<0.02	χ^2=32.5, df=5, p<0.00
Adv. physiological state				
Adverse mental states	χ^2=7.55, df=2, p<0.02		χ^2=25.7, df=8, p<0.00	χ^2=18.3, df=5, p<0.00
Violations				
Perceptual errors				χ^2=12.5, df=5, p<0.02
Skilled-based errors		χ^2=17.1 ,df=8, p<0.02	χ^2=63.6, df=8, p<0.00	χ^2=18.1, df=5, p<0.00
Decision errors			χ^2=35.7 df=8, p<0.00	χ^2=11.7, df=5, p<0.03

Effect of Phase of Flight

At the level of ‘unsafe acts of operators’, the associations of flight phase with ‘decision errors’ and ‘skilled-based errors’ were significant. The flight phase of ‘landing’ was over-represented in these two categories of accidents. At the level of ‘precondition for unsafe acts’, the association of flight phase with ‘adverse mental states’ was significant, as was the association of flight phase with ‘physical/mental limitations’ and with ‘crew resource management’. The flight phase of ‘operational area’ was over-represented in these three categories of accidents. At the level of ‘unsafe supervision’, the association of flight stages with ‘inadequate supervision’ was significant. The flight phase of ‘landing’ was over-represented in the category of ‘inadequate supervision’ of accidents. At the level of ‘organizational influences’, there was no significant association between flight phase and any category within the HFACS framework (see table 2).

Effect of Pilot’s Rank

At the level of ‘unsafe acts of operators’, the association of a pilot’s rank with ‘decision errors’ was significant, as was the association of a pilot’s rank with ‘skill-based errors’ and with ‘perceptual errors’. The rank of ‘lieutenant’ was over-represented in these three categories of accidents. At the level of ‘preconditions for unsafe acts’, the associations of a pilot’s rank with ‘adverse mental states’, ‘physical/mental limitation’ and the ‘physical environment’ were significant. The rank of ‘lieutenant’ was over-represented in categories of ‘adverse mental states’ and ‘physical/mental limitation’ of accidents. However, the rank of ‘lieut. colonel above’ was over-represented in the category of ‘physical environment’ of accidents. At the level of ‘unsafe supervision’, the association of a pilot’s rank with ‘inadequate supervision’ was significant. The rank of ‘cadet’ was over-represented in the category of ‘inadequate supervision’ of accidents. At the level of ‘organizational influences’, the association of a pilot’s rank with ‘organizational process’ was also significant. The rank of ‘cadet’ was over-represented in the category of ‘organizational process’ of accidents (see table 2).

Discussion

The category of ‘decision-errors’ at the level of ‘unsafe acts of operators’ has a significant association with flight phases and rank of pilots. However, it is important to keep in mind that the higher levels affect the next downward level in HFACS framework. It means that decision errors may be affected by ‘precondition for unsafe acts’, ‘unsafe supervisory’, and ‘organizational influences’. This is particularly true of the category of ‘unsafe supervision’ at level-3 of the HFACS. This is one of the key factors, for it not only affects the ‘decision errors’ of pilots, but it also has a significant association with the type of aircraft, mission, flight phase, and rank of pilots (table 2). To precisely identify training needs of ADM, it is necessary to look further into the factors underlying decision errors by applying he HFACS framework.

Although the results showed that fighters had highest frequency of accidents (342), followed by training aircraft (111) and cargo aircraft (56), further analysis found that the training aircraft were significantly associated with ‘adverse mental states’, ‘personal readiness’, ‘inadequate supervision’ and ‘organizational process’. The training aircraft have the highest usage in the Air Force, hence there is time pressure for maintenance, checking processes for airworthiness oversight, and instructor pilots may not have time to provide enough training/supervision. Training aircraft are operated by novice pilots who may not be ready for solo. Cargo aircraft were significantly associated with ‘CRM’ because these types were operating by multi-crew, therefore, CRM was more relevant for crew to perform their tasks than in a one-seat fighter. Fighters were generally under-represented in the HFACS categories. The possible explanation this was that fighter pilots were mature pilots who performed the most demanding tasks in all-weathers, such as interception and air combat tactics. As a result, they were aware of the risk and they were experienced and with a prudent attitude.

There was a significant association between missions and the HFACS framework in three categories: ‘skill-based errors’, ‘personal readiness’, and ‘inadequate supervision’. Further analysis found the task of ‘close pattern’ was over-represented in these three categories of accidents. The possible explanation was ‘close pattern’ practicing of basic take-off and landing skills, was designed for training the novice pilots to operate the aircraft safely. As the pilots were novices with limited experience and operating skills, if the instructor pilots did not provide proper training/supervision, sending a novice solo when he was not ready or had not developed the psychomotor skills, may have resulted in the above three HFACS categories being significant when related to mission of ‘close pattern’.

There was a significant association between flight phase and HFACS framework in six categories. At the level of ‘unsafe acts of operators’, ‘decision errors’ and ‘skill-based errors’ were significantly associated with ‘landing’. In the landing phase, precise psychomotor skills are required to control the aircraft and occasionally instant decisions and responses are needed. At the level of ‘preconditions for unsafe acts’ the categories of, ‘adverse mental states’, ‘physical/mental limitation’, and ‘crew resource management’ were significantly associated with the phase of flying in the ‘operational area’. The possible explanation was that military tactical training such as air combat tactics or low altitude tactics with high physical and mental requirement on the pilots all occur at this stage. Pilots needed to pay more attention to the cognitive demands while flying in the ‘operational area’. They are required to be in a heightened mental state to allow for quick analysis of the dynamic situation to be made followed by swift responses while under time pressure. They also need to have good crew resource management skills to deal with emergent risks and set the priorities for safety issues. At the level of ‘unsafe supervisions’, ‘inadequate supervision’ was significantly associated with ‘landing’. This was perhaps due to the instructors in the MOB not providing enough supervision, providing inappropriate instruction for landing, or back seat instructor pilots failing to provide suitable training for trainees.

The pilot’s rank was related to flying experience. Senior officers normally have more flying hours than junior officers. The rank of ‘cadet’ was significantly over-represented at the categories of ‘organizational process’ and ‘inadequate supervision’. It was perhaps the junior cadet pilots lack of experience and competence to deal with high levels of supervisions and organizational influences, therefore, they were very vulnerable. The rank of ‘lieutenant’ was significantly associated with ‘decision errors’, ‘skill-based errors’, ‘perceptual errors’, ‘adverse mental states’, and ‘physical/mental limitation’. Pilots with the rank of ‘lieutenant’ were the novice pilots (between 200 and 500 flying hours), and at the beginning stage of conversion from training aircraft (AT-3) to fighters (F-16/M-2000/IDF). During this conversion period, it was the tendency of pilots toward having a higher accident rate. The rank of ‘lieutenant colonel (above)’ was significantly associated with ‘physical environment’. The explanation probably that it was only experienced pilots whom were believed to have the ability and the confidence to take the risky tasks in adverse weather or over difficult terrain conditions, so the tasks in an adverse physical environment were assigned to pilots with the rank of lieutenant colonel (and above).

Conclusion

For 25 years, the importance of aeronautical decision-making (ADM) has been recognized as critical to the safe operation of aircraft, as well as accidents avoidance (Jensen & Hunter, 2002). Dekker (2001) described that human errors is systemically connected to the tools, tasks, and operational and organizational environment of operators, it is important to clarify the role of decision errors in pilot’s tools, tasks, experience, and operating environment in military aviation in order to develop effective ADM training programs for military pilots. This research finds pilots at the rank of ‘cadet’ (experience) flying ‘training aircraft’ (tools) practicing ‘close pattern’ (missions) during the ‘landing phase’ (an aspect of the working environment) were likely to be involved in an accident. 'Decision-errors' also had a significant association with the landing phase and lieutenant pilots. However, there are many factors at the upper levels of HFACS framework that will also affect pilots making decisions. It is important to understanding that junior pilots are very vulnerable to the decisions and supervisory practices of senior management

References

Branson, R.K., Rayner, G.T., Cox, J.L., Furman, J.P., King, F.J., and Hannum, W.H. (1975), *Interservice Procedures for Instructional Systems Development*, Report no. N613.39-73-C-O150, Naval Education and Training Command, Florida, U.S.A.

Dekker, S. (2001), 'The Re-invention of Human Error', *Human Factors and Aerospace Safety*, Vol. 1, No. 3, pp. 247-266.

Diehl, A. (1991), 'The Effectiveness of Training Programs for Preventing Aircrew Error ', in Jensen, R.S. (Editor), *Sixth International Symposium on Aviation Psychology*, Vol. 2, Columbus, Ohio, U.S.A., The Ohio State University, pp. 640-655.

Jensen, R. and Benel, R. (1977), *Judgment Evaluation and Instruction in Civil Pilot Training*, Federal Aviation Administration, Washington, D.C.

Jensen, R. and Hunter, D. (2002), *General Aviation Aeronautical Decision-making*, FAA, Washington, D.C.

Landis, J.R. and Koch, G.G. (1977), 'The Measurment of Observer Agreement for Categorical Data', *Biometrics*, Vol. 33, pp. 159-174.

Orasanu, J. (1993), 'Decision Making in the Cockpit', in Wiener, E.L., Kanki, B.G., and Helmreich, R.L. (Editor), *Cockpit Resource Management*, Academic Press, San Diego,

Orasanu, J. and Connolly, T. (1993), 'The Reinvention of Decision Making', in Klein, G.A., Orasanu, J., Calderwood, R., and Zsambok, C.E. (Editor), *Decision Making in Action: Models and Methods*, Ablex, Norwood, New Jersey, pp. 3-20.

Patrick, J. (2003), 'Training', in Tsang, P.S. and Vidulich, M.A. (Editor), *Principles and Practice of Aviation Psychology: Human Factors in Transportation*, Lawrence Erlbaum, New Jersey, pp. 397-434.

Reason, J. (1990), *Human Error*, Cambridge University, New York.

Reason, J. (1997), *Managing the Risks of Organizational Accidents*, Ashgate, Aldershot, England.

Shappell, S.A. and Wiegmann, D.A. (2001), 'Applying Reason: the human factors analysis and classification system (HFACS)', *Human Factors and Aerospace Safety*, Vol. 1, No. 1, pp. 59-86.

Shappell, S.A. and Wiegmann, D.A. (2003), *A Human Error Analysis of General Aviation Controlled Flight Into Terrain Accidents Occurring Between 1990-1998*, Report no. DOT/FAA/AM-03/4, Federal Aviation Administration, Washington, DC.

Shappell, S.A. and Wiegmann, D.A. (2004), 'HFACS Analysis of Military and Civilian Aviation Accidents: A North American Comparison', *International Society of Air Safety Investigators*, Australia, Queensland, pp. 2-8.

Wiegmann, D.A. and Shappell, S.A. (1997), 'Human Factors Analysis of Postaccident Data: Applying Theoretical Taxonomies of Human Error', *The International Journal of Aviation Psychology*, Vol. 7, No. 1, pp. 67-81.

Wiegmann, D.A. and Shappell, S.A. (2001a), 'Applying the Human Factors Analysis and Classification System to the Analysis of Commercial Aviation Accident Data', in Jensen, R.S. (Editor), *Eleventh International Symposium on Aviation Psychology*, Columbus, Ohio, The Ohio State University,

Wiegmann, D.A. and Shappell, S.A. (2001b), 'Human Error Analysis of Commercial Aviation Accidents: Application of the Human Factors Analysis and Classification System', *Aviation, Space, and Environmental Medicine*, Vol. 72, No. 11, pp. 1006-1016.

Wiegmann, D.A. and Shappell, S.A. (2001c), 'Human Error Perspectives in Aviation', *The International Journal of Aviation Psychology*, Vol. 11, No. 4, pp. 341-357.

Wiegmann, D.A. and Shappell, S.A. (2003), *A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System*, Ashgate, Aldershot, England.

MAINTAINING AIRCRAFT ORIENTATION AWARENESS WITH AUDIO DISPLAYS

Kristen K. Liggett, Ph.D.
Human Effectiveness Directorate
Wright-Patterson Air Force Base, Ohio

Peter Venero
General Dynamics
Dayton, Ohio

Martin N. Anesgart, Ph.D.
Human Effectiveness Directorate
Wright-Patterson Air Force Base, Ohio

This study was conducted to determine an appropriate task with which to test alternative orientation display formats, and to test a preliminary set of audio orientation symbology sets. Participants were required to perform three tasks simultaneously. The first task was a visual search (target designation) task. The second task was a radar monitoring task. Both of these tasks were performed on a head-down display. The third task consisted of monitoring aircraft orientation on a head-up display. The third task employed the study's one independent variable – orientation symbology sets. When performing the aircraft orientation task, orientation was displayed in three ways: visual only, visual plus discrete audio orientation information, and visual plus continuous audio orientation information. Performance measures on all three tasks were collected. Results showed that participants responded more quickly to changes in aircraft orientation with the presence of discrete audio orientation information. Lessons learned about the tasks chosen for this study and the audio display symbology sets are discussed.

Introduction

Pilots are required to perform many complex tasks during a mission. Obviously, one of the most important tasks is flying the aircraft, but when multiple tasks are cognitively and visually intensive, pilots can unintentionally lose track of the attitude of the aircraft. Primary flight information is continually presented on a visual display in the cockpit (and sometimes on multiple displays if both a head-down and head-up display are used), and pilots can also obtain visual orientation information from the out-the-window scene given good weather. However, pilots sometimes rely on (often erroneous) vestibular and proprioceptive cues to maintain orientation when performing other visually-intense tasks. When this happens, pilots can easily fall victim to spatial disorientation (SD) (Gillingham, 1992).

SD is defined as "failure to sense correctly the position, motion, or attitude of the aircraft or the pilot within the fixed coordinate system provided by the surface of the Earth and the gravitational vertical" (Previc and Ercoline, 2004, pp. 552). SD is most commonly described as two different types. Type I SD is called unrecognized SD and occurs when pilots are unaware that their perceived orientation is incorrect or different from their actual orientation. This often happens when aircraft undergo sub-threshold movements, causing pilots to perceive their attitude as straight and level when, in fact, they are banking. Type II SD is called recognized SD and occurs when pilots are aware that there is a mismatch between their perceived orientation and their actual orientation as displayed by the flight instruments or the real world. Statistics show that the majority of accidents attributed to SD are caused by Type I, or unrecognized SD. For example, a USAF study reviewing SD mishaps from 1989-1991 showed that 100% of these accidents were attributed to Type I SD (Lyons, Ercoline, Freeman, and Gillingham, 1994). Therefore, the primary goal of this research was to find ways to decrease the occurrence of unrecognized SD by helping pilots maintain orientation awareness throughout the entire mission. Currently, attitude information is primarily acquired visually. But as previously described, the visual channel often becomes overloaded and pilots' attention can become captured by a particular display or task (Foyle, McCann, Sanford, and Schwirzke, 1993; Weintraub and Ensing, 1992). The challenge is determining how one can prevent pilots from losing track of their orientation information when their visual channel is overloaded?

Because the majority of tasks in the cockpit rely on visual resources, audio displays are becoming more popular in the cockpit. Traditional audio displays are basic warnings used to alert pilots when a dangerous situation has arisen. However, audio displays are capable of providing additional information that might help resolve the Type I SD problem. Wickens'

Multiple Resource Model (1984) suggests that if one uses different resources to perform multiple tasks, the tasks can be performed more effectively than if all of the tasks required the same resources. Along those lines, using audio symbology to present attitude information was investigated in this study.

Objective

The objectives of this study were twofold. The first objective was to determine if the task chosen for this study was challenging enough to induce Type I SD when just visual orientation symbology was presented in the cockpit. The second objective was to test the "goodness" of different audio symbology sets for providing additional orientation information.

Method

Participants

Five males and one female participated in this study. Participants were office workers with no piloting experience. The average age of the participants was 27.5 years.

Apparatus

Evaluation cockpit. A fixed-based single-seat generic fighter cockpit simulator was used for this evaluation (Figure 1). It contained a side-mounted, limited-displacement stick with an F-15 grip, and F-15E throttles. The head-down display formats were portrayed on a single 21" x 16" Matsushita color monitor. A BARCO Retrographics 801 system supported the out-the-window scene, providing a 40° horizontal by 30° vertical field of view.

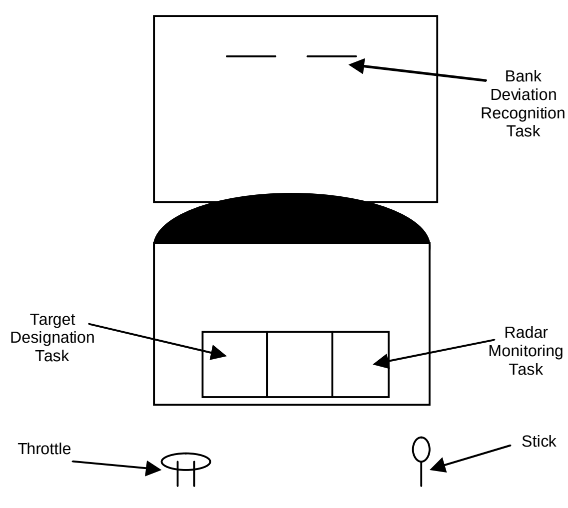


Figure 1. *Evaluation Cockpit*

3-D Localized Audio System. NASA's Sound LABoratory (SLAB) Version 5.3.0 (http://human-factors.arc.nasa.gov/SLAB) (Miller, 2002) generated and presented the 3-D localized audio symbology. SLAB is a software-based real-time virtual acoustic environment rendering system. The software was hosted on a PC and allowed for the specification of position (azimuth and elevation) and volume of the audio input. The third dimension, range, remained fixed for this study. SLAB interfaced with a head tracker to receive head orientation information and modify the location of the sound so the location of it appeared stationary. The audio symbology was presented via Panasonic headsets, which were worn by the participants during the study.

Head Tracker. An Ascension Flock-of Birds 6-D Multi-Receiver/Transmitter Tracking Device was attached to the participant's headset to measure head position coordinates and orientation angles. This information was sent to the 3-D audio system to ensure that the 3-D audio tones were properly correlated with the participant's head position.

Cockpit Tasks

Participants were required to perform three tasks simultaneously. Two of the tasks were conducted on the head-down display; the third was conducted on the head-up display.

Target Designation Task. This task was employed on the left portion of the head-down display (Figure 1). The goal of this task was to find the target symbol (diamond) as fast as possible. Also present with the target symbol were 252 distracter symbols; 84 boxes, 84 upright triangles, and 84 upside-down triangles. Figure 2 shows a sample screen of the target designation task. To select a target symbol, the participant slewed a button on the throttle until the cursor on the screen overlaid the target symbol. Then the participant pressed a button on the control stick to designate the target. As soon as the participant designated the correct target, a new screen appeared. Participants designated as many targets as they could before the trial was completed. Trial length was dependant on the bank deviation task.

Radar Monitoring Task. This task was also employed on the head-down display to the right of the target designation task (Figure 1). The goal of the radar monitoring task was to keep the strength of the radar at an optimal level. This was achieved by keeping a status bar as close to the center mark (0) as possible. Figure 3 shows the radar monitoring task display.

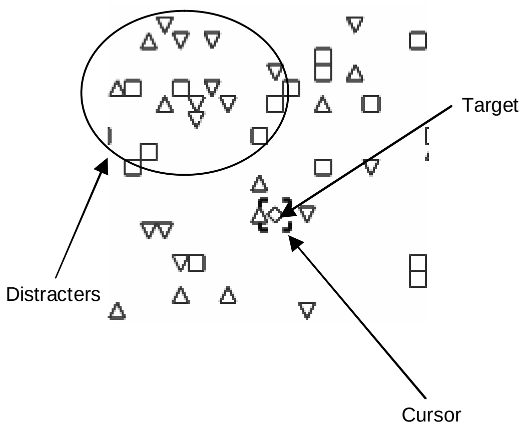


Figure 2. *Target Designation Task*

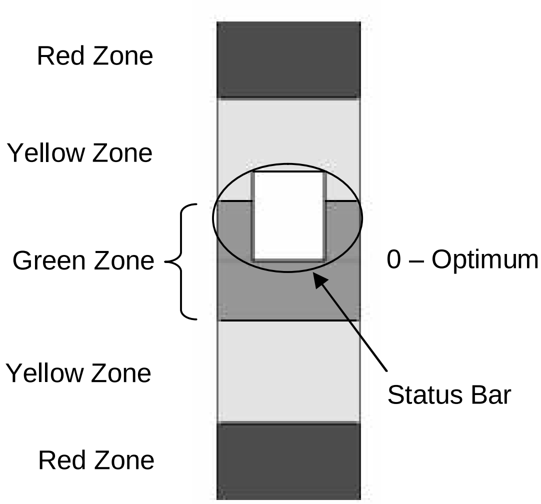


Figure 3. *Radar Monitoring Task*

The status bar was driven by a function of four sine waves. The participant had to press a switch on the control stick in the down position to get the status bar to move down, and press the same switch in the up position to get the bar to move up. Each switch hit would move the bar a discrete amount, and then the status bar would resume moving according to the function of sine waves.

Bank Deviation Recognition Task. This task was employed on the head-up display, which contained a blue background and a green line (Figure 1). Although the participants were not actually flying an aircraft, the line represented aircraft bank. The purpose of the task was to recognize and correct any bank deviation that occurred. Once a bank deviation was recognized, the participant had to move the control stick in the direction opposite of the bank angle to correct it. For example, if the bank indicator rolled to the left, the participant had to move the control stick to the right to level out the bank deviation. The bank indicator moved at a rate of 10º/s to a maximum of 30º bank, moving to and from wings level. If the participant corrected the bank deviation before the bank indicator reached the full 30º of bank, the bank indicator immediately moved back to 0º bank. The bank angle deviations were presented to the participants at random times. There was a total of 25 right bank deviations and a total of 25 left bank deviations per task. The direction of the bank deviations were presented randomly. The time between bank deviations was random and varied between 0 and 6 seconds.

Audio Symbology

To determine the effects of the addition of audio symbology for maintaining orientation awareness, three different conditions were tested during this study. The first condition was a visual only task in which participants could observe bank deviations only by looking at the visual head-up display. The second condition was a visual plus discrete audio orientation symbology task. In this condition, in addition to the visual head-up display, a discrete audio pattern was activated when the bank indicator deviated from 0º in either direction. Once the audio pattern became active, a 100 ms white noise sound source would pulse at 0.5 Hz directly in front of the participant. The audio display did not stop pulsing until the bank indicator returned to 0º bank. The third condition was a visual plus continuous audio orientation symbology task. In this condition, in addition to the visual head-up display, a continuous audio pattern was active at all times. When bank was 0º, the sound source (100 ms white noise) was located directly in front of the participant. When the bank deviated to the left, the sound source moved to the left at a fixed rate of 30º/s with a maximum displacement of 90º to the left, and vice versa for the right. The fixed rate at which the sound source moved was three times as fast as the bank indicator movement.

Independent and Dependent Measures

There was one treatment variable, Audio Orientation Symbology, with three levels measuring the effect of the addition of audio symbology on the bank deviation task. The three levels were visual symbology only, visual symbology plus discrete audio symbology, and visual symbology plus continuous audio symbology.

Dependant measures were collected for all three cockpit tasks. The dependent measures collected for the target designation task were number of correct targets designated and average search time for a target. For the radar monitoring task, root mean square (RMS) errors from the optimum position was the dependent measure. These measures tested the first objective, which was how much effort the participants were giving to the head-down tasks, and in turn, getting mentally loaded. For the bank deviation recognition task, the dependent measure was average time to react to bank deviations. This measure tested the second objective, which was the effect of adding the audio symbology to the visual orientation symbology for recognizing bank deviations.

It was hypothesized that the audio symbology would enhance performance on the bank deviation recognition task and that the continuous symbology would be the more helpful of the two audio symbology sets. It was also hypothesized that by enhancing performance on the bank deviation task with the addition of audio, performance on the other two tasks would increase due to the lessening of the visual load that would occur when the bank deviation task was augmented by the audio symbology.

Experimental Design

The study had a completely within-subjects design using three levels of one treatment variable – Audio Orientation Symbology. In an effort to control practice effects, complete counterbalancing of the three levels was used.

Procedure

Subjects were first given a standardized briefing on safety and test procedures. Next, the three tasks were explained and training ensued. First, training on the target designation task was conducted. This was broken down into three levels of difficulty. The easiest level consisted of finding the target symbol among 252 box distracter symbols. The next level of difficulty included the target symbol with 126 box distracter symbols and 126 upright triangle (with the point at the top) distracter symbols. The final level of difficulty, and the one used for data collection, included the target symbol, 84 boxes, 84 upright triangles and 84 upside-down triangles (balancing on their point). The participants were given two practice trials at each level of difficultly. Then practice proceeded with the target designation task and the radar monitoring task simultaneously. Finally, training on all three tasks occurred in which participants were given 15 left and 15 right bank deviations each. Participants were instructed to give equal priority to all three tasks. The data collection consisted of three trials that were the same as the practice trials save one detail – the collection trials contained 25 left and 25 right bank deviations.

Results

Determining effects of the Audio Orientation Symbology condition on participant's ability to recognize and correct bank deviations, while at the same time performing a target designation task and a radar monitoring task required a sophisticated statistical procedure called *repeated measures multivariate analysis of variance*. This procedure permits joint testing of multiple dependent variables.

Proper use of the multivariate procedure necessitated correlating the four dependent measures beforehand and creating models based on these correlations. Table 1 shows the correlation matrix.

Table 1. *Correlation Matrix*

	Reaction Time (Bank Task)	Number Targets (Target Designtn Task)	Search Time (Target Designtn Task)	RMS Errors (Radar Task)
Reaction Time	1	0.048	0.356**	0.316*
Number Targets		1	-0.767**	-0.35**
Search Time			1	0.527**
RMS Errors				1

* p<0.05
**p<0.01

The pattern of correlations required testing two models –Model 1: Reaction Time, Search Time and RMS Errors; Model 2: Number of Targets designated, Search Time and RMS Errors.

For both models, given an N of only 6 participants, the lack of residual (or error) degrees of freedom precluded the robust (to violation of statistical assumptions) omnibus multivariate tests. Fortunately each of the dependent measures in the two models did not violate sphericity for the audio condition, thus enabling "averaged F" tests of the models. In Model 1, there was a significant effect for the Audio Orientation Symbology condition ($F(6,18)=3.071$, p=.030). The strength of the effect (based on Pillai's Trace) was moderate with an eta^2 (partial) of 0.506. Also in Model 2, there was a significant effect for the Audio Orientation Symbology condition ($F(6,18)=2.988$, $p=.033$). The strength of the effect was again moderate and based on Pillai's Trace with an eta^2 (partial) of 0.499. Roy's Largest Root showed stronger effects, an eta^2 (partial) of 0.773 for Model 1 and an eta^2 (partial) of 0.788 for Model 2.

The significant averaged F tests allowed separate tests of each dependent measure in the two models. These univariate tests revealed only Reaction Time as significant for Audio Orientation Symbology condition ($F(2,10)=15.933$, $p=.001$), with a strong effect, eta^2 (partial) of .761. Figure 4 shows the mean reaction time for each audio condition.

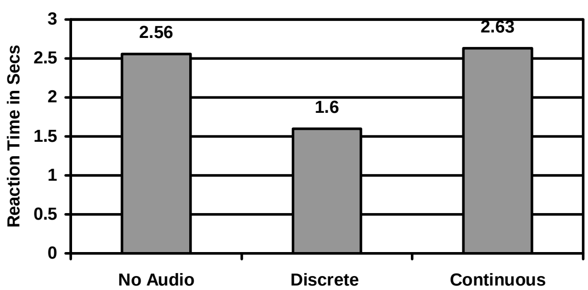


Figure 4. *Raw Means for Audio Orientation Symbology Conditions*

Two further tests revealed which differences among the audio condition levels were strongest: a test of within-subjects contrasts, and Bonferroni paired comparisons. The three levels of the audio symbology condition made two contrasts available. The first compared the discrete mode to continuous, while the second compared the no audio condition to the average of the other two audio conditions. Both contrasts reached significance ($F(1,5)=20.237$, $p=.006$; eta^2 (partial) of 0.802 and $F(1,5)=8.678$, $p=.032$; eta^2 (partial) of 0.634, respectively). The significance found for the second contrast was due to the difference between the discrete mode and no audio as the Bonferroni comparisons attest. The discrete mode was significantly ($p<.05$) on average 1.028 seconds faster (with a standard error of 0.228 seconds, $p=.019$) in Reaction Time than the continuous mode. It was also on average significantly faster ($p<.01$) than the no audio mode, 0.963 seconds (with a standard error of 0.180 seconds, $p=.009$). There was, however, no significant difference ($p>.05$) in Reaction Time between the continuous and the no audio modes (a mean difference of 0.065, standard error of 0.200, $p=1$). Note that mean differences among the Audio Orientation Symbology conditions are slightly at variance with those shown in Figure 4; the Bonferroni procedure bases its comparisons on the estimated marginal means from the models.

Discussion

Results showed that the discrete audio orientation symbology significantly helped participants notice bank deviations more quickly than with the continuous audio orientation symbology or with no audio symbology at all. This is an interesting finding given that the continuous audio symbology provided more information to the participants in terms of the direction of the bank deviation. Recall that the discrete audio symbology sounded in the same manner whenever deviations from 0º bank occurred. It basically provided an audio alarm of bank deviations. The continuous audio symbology sounded when deviations occurred and were presented in the direction of the bank deviations.

Informal questioning of the participants revealed that reaction time on the continuous audio may have been slower because they were initially unable to tell which direction the tone was moving. In other words, they had to wait until they could accurately localize the tone before they could respond, which delayed their reaction time. The disadvantage to this strategy is that it took them longer to respond to the bank deviations. However, the advantage to this strategy is that, once they could determine the direction of the continuous audio, they would respond without taking their visual attention away from the head-down tasks. Therefore, if an adjustment to the continuous audio symbology to allow for quicker initial position detection is possible, this may transition the orientation awareness task to a purely audio task verses yet another visual task.

Since participants were delaying their reaction time to the bank deviation task in the continuous audio condition so they could keep their visual attention focused on the head-down tasks, one would expect to see a performance enhancement in terms of the dependent measures for the other two tasks during the continuous audio symbology condition. This, however, was not the case. Overall results suggest that neither of the audio symbology enhancements had an impact on a participant's head-down work demands per se as shown by the non-significant effects for the audio conditions in terms of the target designation task dependent measures (search time and number of targets designated) and the radar monitoring task dependent measure (RMS errors).

This may be attributed to the fact that, although participants were not using their *visual resources* to perform the bank deviation task while using the continuous audio symbology, they were still required to *cognitively* attend to the task, which competed with head-down task resources.

This argument of cognitive attentional requirements for the bank deviation task versus the head-down tasks holds true for the discrete audio symbology as well. Although results showed a significant decrease in reaction time for the bank deviation task when discrete audio orientation symbology was present, participants were required to go head-up to determine the direction of bank, make the appropriate control stick input, and return to the head-down tasks. Therefore, even though participants performed the bank deviation task faster, it took cognitive resources away from the head-down tasks, and performance enhancements to the head-down task were not found.

It seems certain that the head-down tasks are challenging for the participants to accomplish, and a ceiling effect may be occurring. In other words, regardless of the type of audio symbology used in the bank indication task, the head-down tasks alone are difficult enough to keep the participants busy. Freeing up a small amount of resources as in the discrete and continuous audio symbology conditions, was not enough to show a significant performance enhancement on the head-down tasks. The advantage of using the audio to help with the bank deviation task is evident in the bank deviation task, but not strong enough (yet) to carry over benefits in performing the other head-down tasks. Adjustments to the audio orientation symbology may show these benefits in future studies.

Interestingly, participant performance during the no audio condition of the bank deviation task was just as good as with the continuous audio symbology. This may be attributed to the fact that participants were told to give equal priority to all three tasks. When there was no audio augmentation, participants relied more heavily on a good visual cross-check pattern to detect bank deviations. In any case, it appears that the tasks chosen for this study were challenging enough to provide a good protocol for testing countermeasures for Type I SD in future studies.

Conclusions

This study was successful in that it adequately tested the study objectives. Although the hypotheses were not proven, lessons learned from this study will be leveraged in future studies which will continue to look at ways of reducing Type I SD and augmenting audio orientation symbology to help combat this problem.

Acknowledgements

The authors would like to thank Tony Ayala, Lee Berger, and Sam Longo for their expertise in software development and hardware integration. Jay Hudepohl, Matt Mantei, and Doug Zimmer are acknowledged for their pilot expertise and valuable time during study development and validation.

References

Foyle, D. C., McCann, R. S., Sanford, B. D., and Schwirzke, M. F. J. (1993). Attentional effects with superimposed symbology: Implications for head-up displays. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (pp. 1340-1344). Santa Monica, CA: HFES.

Gillingham, K. K. (1992). The spatial disorientation problem in the United States Air Force. In *Journal of Vestibular Research, Vol 2*, pp. 297-306.

Lyons, T. J., Ercoline, W. R., Freeman, J. E., and Gillingham, K. K. (1994). Classification problems of U.S. Air Force spatial disorientation accidents, 1989-1991. In *Aviation, Space, and Environmental Medicine, Vol 65,* pp. 147-152.

Miller, J.D. and Wenzel, E.M. (2002). Recent developments in SLAB: A software-based system for interactive spatial sound synthesis. In *Proceedings of the International Conference on Auditory Display* (pp. 403-408). Kyoto, Japan: ICAD.

Previc, F. H, and Ercoline, W. R. (2004). *Spatial Disorientation in Aviation.* Reston, VA: American Institute of Aeronautics and Astronautics.

Weintraub, D. J., and Ensing, M. J. (1992). *Human Factors Issues in Head-Up Display Design: The Book of HUD.* Alexandria, VA: Defense Logistics Agency.

Wickens, C. D. (1984). *Engineering Psychology and Human Performance*. Columbus, OH: Charles Merrill.

WORKSPACE VISUALIZATION FOR PLANNING OF AIR OPERATIONS

Gavan Lintern
General Dynamics, Advanced Information Systems
Dayton, Ohio 45431-1289
glintern@earthlink.net

Information overload has become a critical challenge within military operations. However, the problem is not so much one of too much information but of abundant information that is poorly organized and poorly represented. Here I describe a prototype information-action workspace, sometimes referred to as a knowledge visualization, to resolve this issue. Development proceeded through a systematic design sequence of cognitive analysis, knowledge representation and workspace design. The cognitive analysis focused on the specific information needed to support military planning and judgment. The workspace was structured in terms of dimensions of functional abstraction and functional decomposition; dimensions that are thought to characterize the fundamental structure of cognitive work. The products of a Cognitive Work Analysis were integrated with insights drawn from operational and scientific literature to develop a prototype workspace. Here I outline some of the features of the prototype workspace.

Information Management

Information management has emerged as a significant contemporary challenge in modern warfare. The advantage now goes not to those with the more potent weaponry but to those with the more effective information system. Commanders and planners can access a huge amount of information. Where that information is about current status and progress of events, it becomes available with unprecedented speed. It has become available in different forms, at different levels of abstraction and from multiple and diverse sources.

Nevertheless, this information is poorly organized. It is available from diverse sources and in fragments, which leaves a commander or planner with the challenge of searching the information space to find, distinguish, summarize, integrate and understand the meaningful elements that can make a difference throughout the execution of a battle plan. That is both an onerous and a difficult task. In a high-tempo, high-stress environment it will often be an impossible one.

Information has always been central to military success but in the modern military, the importance of its role is increasing. Nevertheless, successful action is rarely based on a mass of information; it typically results from decisions in response to key pieces of information that become available at the right time. A timely decision based on a few key observations can turn a potential disaster into a victory. However, to achieve victory, decision-makers must be able to recognize and to act on the opportunities available to them. That requires a well-designed interface; termed in this paper a *virtual information-action workspace*.

Ecological Interface Design

… *the conclusion is unmistakable: if displays of data are to be truthful and revealing, then the design logic of the display must reflect the intellectual logic of the analysis.*
Edward R. Tufte (1997), Visual Explanations, p 53

A central assumption of Ecological Psychology is that the functional needs of an organism necessarily reciprocate the functional structure of that organism's natural world (Reed and Jones, 1982). In accordance with that assumption, Ecological Interface Design results in a virtual world that reciprocates the structural constraints on cognitive work. The term *ecological* is drawn from the field of Ecological Psychology in which the driving interest is the relationship between an organism and its environment (Gibson, 1979).

The key tenets of Ecological Psychology as relevant to interface design are:

- Human action is constrained by the work domain
- Interfaces are mediated environments that can reveal the work constraints
- Information can be depicted in a manner that supports direct perception of those constraints

The approach of Ecological Interface Design is to analyze the work domain, to identify its constraints and to then develop perceptual forms that reveal the constraints directly at the interface.

The general claim driving the work reported here is that the information gathered from the world by technical sensors and human observers enters the planning information system as a fragmented and disorganized set. Some form of human-centered analysis and design must be applied to organize that information and extract its significant meaning. Most forms of human-centered analysis start by addressing cognitive limitations or user preferences. There is typically little explicit concern with the structure of the work domain. Ecological Interface Design starts from the other direction; a consideration first of the structure of the work domain followed by a consideration of how the user might interact with it.

Design Strategy

The design process for building a virtual information-action space to support cognitive work progresses through four distinct stages; knowledge acquisition, knowledge representation, design specification and fabrication. The role of Ecological Interface Design is oriented around the first three of these, leaving fabrication to the applied engineering disciplines.

The principles and procedures of Cognitive Work Analysis and representational forms drawn from the human factors display literature and from work domain publications are used to develop virtual workspace specifications for:

- Information requirements (what information should be displayed)
- Information layout (how information should be organized relative to other information)
- Workspace navigation (the capabilities needed to search for and integrate or associate different information elements)
- Action on the work domain (the form, content and magnitude of transactions between entities)
- Information representation (how information should be represented so workers can rapidly perceive its meaning)

Work Domain Analysis (one stage of Cognitive Work Analysis) was used to specify information requirements and layout for a workspace.

Work Domain Analysis

A Work Domain Analysis results in a knowledge representation termed an Abstraction-Decomposition map. This map catalogues the functional properties of the work domain (objects, resources, constraints, purposes) in a two-dimensional matrix in which the vertical dimension represents levels of abstraction and the horizontal dimension represents varying levels of decomposition.

The upper three levels of abstraction (System Purpose, Values & Priorities, Purpose-Related Functions) identify the intentional (socio-organizational) constraints of the system while the lower two levels (Physical Functions, Physical Properties) identify its physical (causal) constraints. This form of representation can be used to specify the information requirements of a work domain. Each node in the Abstraction-Decomposition map points to information (either directly or indirectly) that must be provided within the workspace. How this can be accomplished has been described in Linegang and Lintern (2003), Lintern, Miller and Baker (2002), Lintern (2002) and Lintern (submitted).

The guiding vision for an information-action workspace is one in which essential information is readily accessible and presented in succinct and meaningful forms. This suggests that there should be summaries of contextually relevant information and readily apparent signs to guide access to it. Evocative visual forms should be used to the extent possible but text information required for support of cognitive work should be summarized and highlighted so that the analyst can converge readily on its essential meaning. In this section, I outline how the results of the analysis as reported by Lintern (submitted) were integrated with selected design concepts to develop a prototype of a virtual information-action workspace.

Workspace Organization

A planning workspace must present a global structure while it provides access to detail; support for the interplay between top-down and bottom-up exploration that characterizes the cognitive activity associated with planning and deciding. The workspace architecture described here follows the single-window, multi-panel format used by Linegang and Lintern (2003) and Lintern, Miller and Baker (2002) for ecological interfaces developed to deal with the cognitive challenges of military command and planning.

A Prototype Workspace

The layout for the workspace is shown in figure 1. Typically, information related to intentional constraints is distributed throughout the panels in a default view of the workspace while information related to physical constraints is brought into view by interrogation within default view panels. Previous work (Lintern, 2002; Lintern, et al, 2002) suggests that the top left panel be allocated to System Purpose and the top right panel to Values and Priorities. The central panel was allocated to a geospatial representation. By this means it was possible to distribute the essential resources for activity within the geospatial area around its periphery as is consistent with the Focus-Periphery Organization Principle (Eggleston, 2002).

The default view of the prototype workspace is shown in figure 2. The top right panel of figure has a Polar Star that depicts normalized parameters associated with Priorities and Values. The top center panel provides access to documents related to System Purposes and Priorities and Values. The cognitive analysis needed to determine the content and style of such documents has not yet been done but these resources are envisioned as succinct summaries of no more than a page or two organized to be relevant to a general context selected via the three-by-three matrix of buttons to the left. The dimensions of this matrix are currently conceptualized as Type of Effect (Physical, Systemic, Psychological) by Level of War (Tactical, Operational, Strategic) as consistent with the results of the analysis (Lintern, submitted).

The Situation Display in the center panel is the primary workspace in which planners or commanders might *drag-and-drop* items from the Allied and Adversary resources panels to the left and right (respectively) and might relocate those resources (as in the old style *sand table*) or interrogate their functional and physical properties. That interrogation could activate more detailed views in the bottom left or right panels.

Information relevant to action within the Situation Display might be assembled in the Problem Work Space (bottom center panel) to explore possibilities for Course of Action (both Allied and Adversary). One of the recurring themes coming out of the analysis was the concern of planners with relationships between allied and adversary capabilities and with the effects of environment on operations. The Problem Work Space of figure 2 is based on a capsule scenario in which a planner is concerned with effects of dust storms on operations. Further exploration would link both allied and adversary capabilities to the information assembled in this panel to examine possible impact of those dust storms on current or potential operations.

Figure 3 depicts how more detailed resources might be brought in to view. The Polar Star for System Purposes shows a problem with one parameter. A depiction of a time history for that variable may be brought into view by clicking on the shortened spoke. This particular format, developed by Tufte (1997), shows status some months in the past, a few days ago, and daily over the last week, with bars showing the limits of normal range. The goal is to remove the problem of understanding what is happening with this variable so that the planner can move quickly into the cognitive problem-solving mode of ascertaining why it is happening (Tufte, 2003).

By interrogating a resource that has been activated in the bottom left panel as a more detailed view, it is possible to bring up more information on that weapons system, in this case a graphic depiction of weather effects on the targeting performance of that system.

Selection of a document icon in the top center panel can open a summary related to Values and Priorities, in this case a summary of Rules of Engagement. The subject matter experts had noted that planners would be familiar with the Rules of Engagement but would occasionally need to check or confirm subtle specifics and may have to do so under time pressure. That forces a scan of a large document; a particularly onerous requirement in a time stressed situation. The pop-up summary, taken from United States Marine Corps (1998), is intended to resolve that problem by having a succinct and pertinent summary at hand.

The top center panel (figure 2) has a video display area and a video library. In addition, a number of photographs are used in the workspace. The inclusion of these items was based on the materiality arguments of Hayles (1999). Nevertheless, these depictions do not yet convey much more than the basic idea. Further cognitive analysis is required to ascertain the character and content of the visual narratives that could satisfy this requirement. It is likely that at least some of these visual narratives will have to be updated frequently (e.g., daily). The source of such resources and the way in which they might be designed to evoke the desired sensitivity to situational events has yet to be determined.

Conclusions and Future Directions

The final product of this research is envisioned as a worktable with an electronic surface on which it will be possible to manipulate computer representations of information structures. It will have a graphical interface that will rely heavily on iconic representation of critical properties. It will have many of the standard tools of graphics programs (e.g. icon libraries, electronic pens, default shapes, connectors) and many of the standard means of computer interaction that permit intuitive and direct selection (touch activation, drag and drop, selection, pointing and linking).

There is considerable cognitive analysis and design work required as yet to achieve the vision of a fully integrated collaborative workspace. As noted in the discussion of the pop-up summary for Rules of Engagement (Figure 3), the requirement for this type of resource was identified in the analysis. Although the summary shown in figure 3 was taken from a military document (United States Marine Corps, 1998), the content and form for a resource such as this should be developed through an analysis and design process similar to the one used to develop the workspace prototype but focused on this particular element. Many other elements of the workspace also demand this sort of effort.

Acknowledgement

This work was completed under contract F33615-98-D-6000, Asymmetric Adversary Analysis for Intelligent Preparation of the Battlespace, to the Air Force Research Laboratory with Robert G. Eggleston as the technical monitor.

References

Eggleston, Robert G. and Whitaker, Randall D. (2002). Work centered support system design: using frames to reduce work complexity. *Proceedings of the 46th Human Factors and Ergonomics Society Annual Meeting.* (pp. 265-269). Santa Monica, CA: Human Factors and Ergonomics Society. [CD-ROM]

Gibson, James J. (1979). *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin.

Hayles, N. Katherine (1999). How we became post human: Virtual bodies in cybernetics, literature, and informatics. Chicago: University of Chicago Press.

Linegang, Michael P & Lintern, Gavan (2003). Multifunction Displays for Optimum Manning: Towards Functional Integration and Cross-Functional Awareness. *Proceedings of the 47th Human Factors and Ergonomics Society Annual Meeting. (pp. 1923-1927).* Santa Monica, CA: Human Factors and Ergonomics Society. [CD-ROM]

Lintern, Gavan (submitted). A Virtual Information-Action Workspace for Military Intelligence Planning. Submitted to the International Journal of Industrial Ergonomics.

Lintern, Gavan (2002). Work Domain Analysis for Distributed Information Spaces. In the *Proceedings of the 2002 International Conference on Human-Computer Interaction in Aeronautics (HCI-Aero 2002).* Menlo Park, CA: American Association for Artificial Intelligence Press, p 94-99.

Lintern, G., Miller, D. and Baker K. (2002). Work Centered Design of a USAF Mission Planning System. *Proceedings of the 46th Human Factors and Ergonomics Society Annual Meeting.* (pp. 531-535). Santa Monica, CA: Human Factors and Ergonomics Society. [CD-ROM]

Reed, Edward and Jones, Rebecca (1982). Reasons for Realism: Selected Essays of James Gibson. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tufte, Edward R. (1997). *Visual Explanations*: *Images and Quantities, Evidence and Narrative.* Cheshire, CT: Graphics Press.

Tufte, Edward R. (2003, August). One Day Course: Edward Tufte, Presenting Data and Information, New York, NY.

United States Marine Corps (1998). *Military Operations on Urbanized Terrain (MOUT).* Marine Corps Warfighting Publication (MCWP) 3-35.3, Washington, DC: Department of the Navy, Headquarters United States Marine Corps.

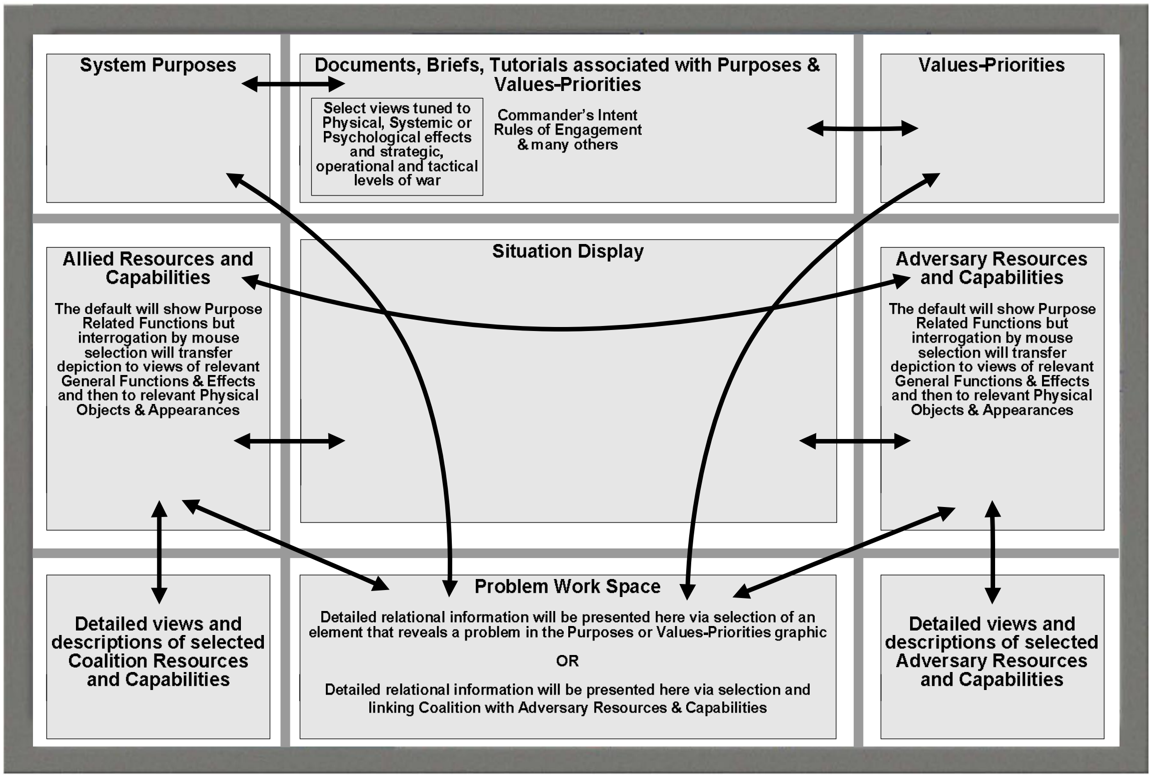


Figure 1. *A distribution of functions within the multi-panel format as derived from Abstraction-Decomposition matrices.*

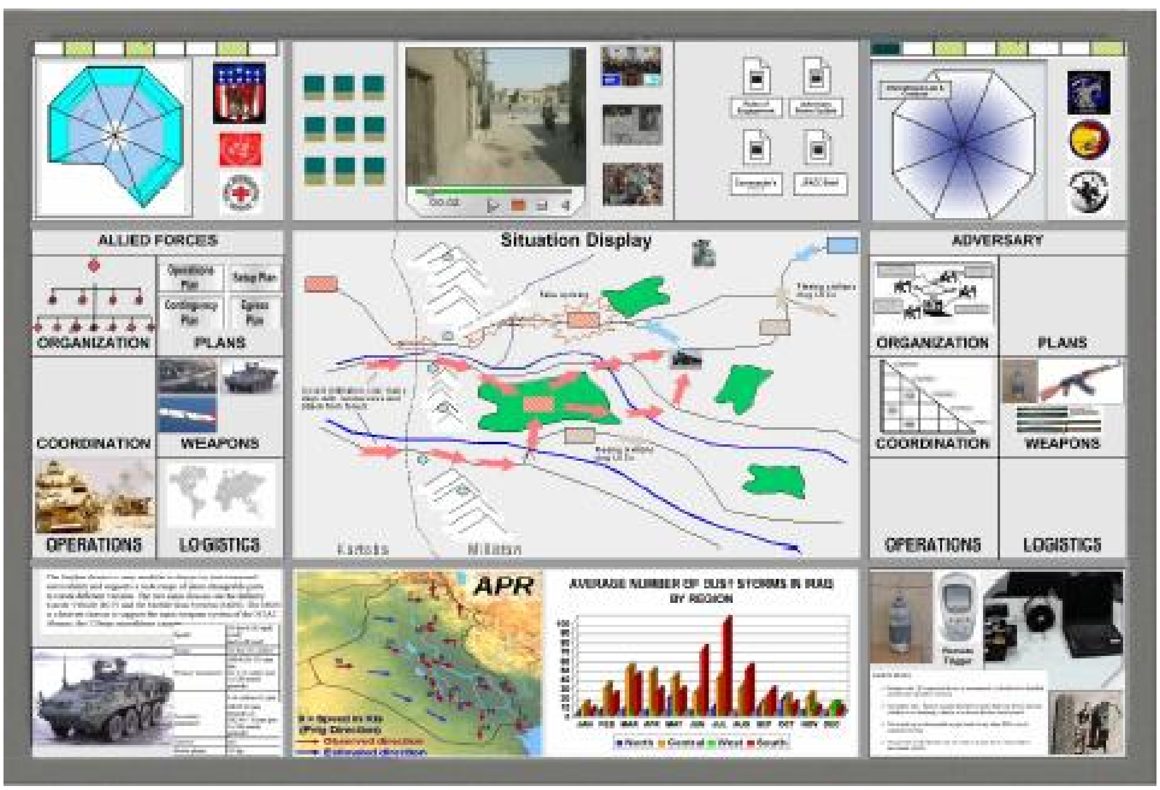


Figure 2: *A depiction of an information-action workspace for Intelligence Preparation of the Battlespace.*

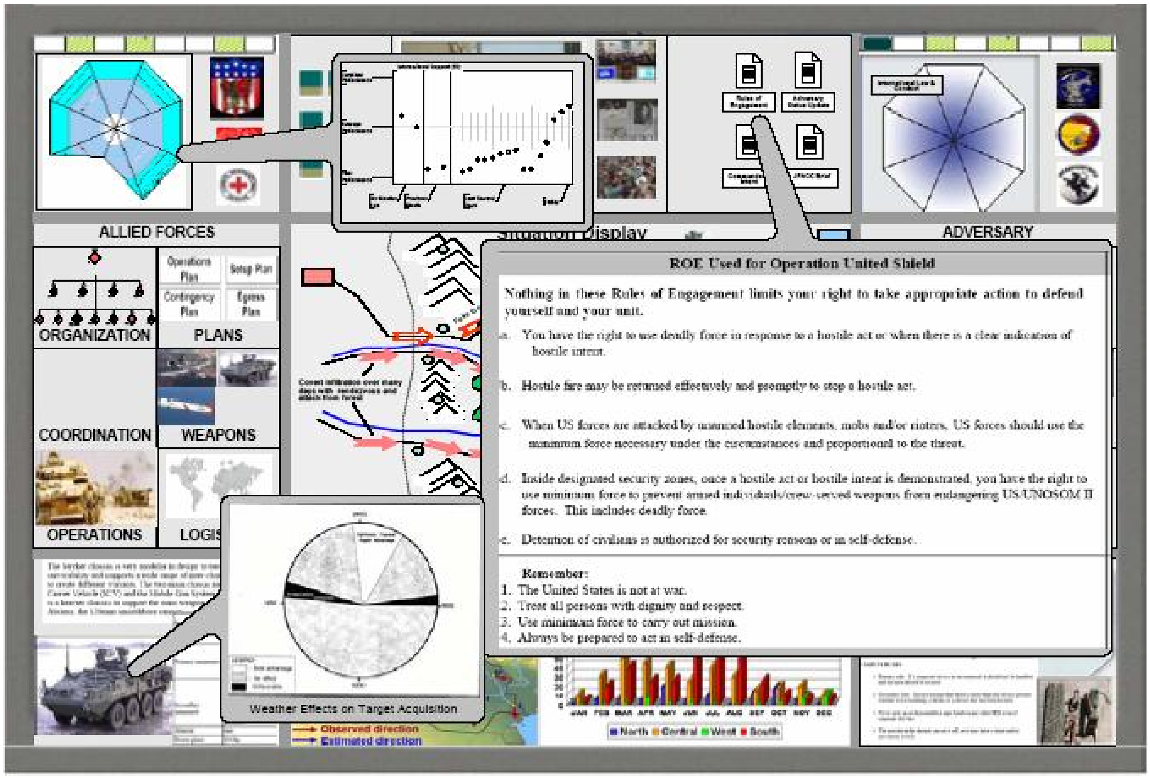


Figure 3. *An illustration of how information resources can be accessed within the information-action workspace depicted in Figure 2.*

THE USE OF INTEGRATED HISTORICAL AND PREDICTIVE DATA TO SUPPORT FLIGHT PLANNING BY AIRLINE DISPATCHERS

Chang Liu, **Philip J. Smith**, and **Charles Billings**
The Ohio State University
Columbus OH

Kapil Sheth, **Shon Grabbe** and **Banavar Sridhar**
NASA Ames Research Center
Moffett Field CA

Airline dispatchers play a critical role in the National Airspace System (NAS), as their flight planning decisions have a direct impact on the efficiency and safety of the resultant traffic flows and on contingency plans to deal with possible events that could arise while enroute. Their decisions also have an important impact on the operating costs for an airline. This paper first discusses the results of two focus groups with airline and military dispatchers that served to identify potential uses by dispatchers of the functionality contained in NASA's Future ATC Concept Evaluation Tool (FACET). This tool uses trajectory modeling to generate predictions of ATC sector loadings and to generate and evaluate alternative routes for an aircraft in terms of potential traffic congestion concerns. The paper then discusses follow-up work based on one of these findings: The potential value of combining data from FACET with historical data about a flight's past performance in order to improve pre-flight planning and flight-following while an aircraft is enroute.

Study 1

An initial study was conducted to identify potential uses of the functions embedded in the Future ATM Concepts Evaluation Tool (FACET) for Airline Operations Centers (AOCs). In addition, the study sought to determine potential enhancements of FACET that might better support the needs of dispatchers and air traffic control coordinators at AOCs.

As part of the study, a total of 19 dispatchers were interviewed. All of them had at least 8 years of dispatch experience. Eight of them also had at least 4 years of experience as an airline ATC coordinator. These dispatchers and ATC coordinators represented experience with dispatching at 5 different airlines. Another of them was a dispatcher for the US Air Force. The participants were introduced to the current capabilities of FACET and asked to consider potential uses and extensions of the functionality contained in FACET for AOCs and the interface design features associated with these functions. Key findings are summarized below.

AOC Tasks that Could Make Use of FACET

FACET was originally designed to support decision-making by FAA traffic flow managers. The dispatchers interviewed identified a number of potential areas where it could be of value to AOCs, however:

- Evaluate traffic constraints along alternative routes for a single flight during preflight planning.
- Identify modifications to a flight plan (route, altitude profile, departure time, speed) that would avoid a traffic constraint.
- Evaluate alternative reroutes contained in ATCSCC reroute advisories in terms of traffic constraints.
- Alert the dispatcher if a flight with an already filed flight plan (whether still pre-departure or enroute) is now predicted to encounter traffic constraints.
- Allow an ATC Coordinator or dispatcher to look at the predicted traffic congestion for specific airspace regions (such as the arrival sectors for an airport).

Predicting Which Flights Will be Moved

One of the key features of FACET is its predictions of air traffic congestion in a sector. Although this type of metric is of use to AOC staff, in many cases the question they really want to answer is how likely it is that this particular flight will be rerouted because of traffic congestion, and if so, what the resultant reroute and airtime is likely to be. Such information would help the dispatcher decide whether to plan a different route or just plan for contingencies if the flight is likely to be tactically moved by ATC (adding extra fuel, etc.).

Incorporating Other Data into FACET

Philosophically, the dispatchers recommended a human-centered approach that treats FACET as one source of data to help the dispatcher make judgments:

"Show them the data and let the person do the probabilistic reasoning."

The dispatchers interviewed indicated that to improve prediction accuracy and help the dispatcher make better judgments, three kinds of data could be integrated into FACET:

- Complete 4-D trajectories based on airline flight plans
- Weather data
- Historical data about the performance of a flight (such as its history of reroutes).

In terms of the use of historical data, the dispatchers noted:

"If you had the ability to show what that specific flight had done on previous days, that could be used in your decision making processing by saying 'okay, this is what happened to me in the last four to five days.'"

"If you had the previous history as to what that flight has done, it would go a long way toward helping you make a decision as to what you are going to do with that flight today. Because if you know if this airplane gets moved 40% of the time, then maybe you would be better off just moving it."

"The first flight is a good predictor of what is going to happen for the rest of the day if nothing major changes. You tend to do the same thing the rest of the day."

Study 2

Based on the recommendation to integrate FACET predictive displays with historical data on flight performance, a set of designs were prepared and then evaluated using a questionnaire. The results are summarized below.

Biographical Data

Fifteen dispatchers were sent a questionnaire about the integrated displays shown in Figures 2-5. All 15 responded. These dispatchers worked for 6 different airlines. Their years of experience dispatching ranged from 8-29 years, with a mean of 17.7 years. All but 3 of them also had experience as ATC Coordinators. These 12 dispatchers had 1-15 years of experience as ATC Coordinators, with a mean of 7.5 years of experience.

Preferred Form of Access

Before discussing details of the design, the dispatchers were asked the following question: Would you prefer to have this information displayed to you for every flight, or only for those flights where the predicted or historical data indicates a potential problem (as an alert)? Please indicate your reasons.

The responses emphasized the need to consider the time pressure often faced by dispatchers when preparing a flight release. Generally speaking, the dispatchers indicated that, although the dispatcher should be able to call up such information about an individual flight, these displays should not be shown for every flight. Instead, the dispatcher should be able to set some parameter(s) that determines when an alert would be generated for a flight, which would then allow the dispatcher to look at the combined predictive/historical data displays for that flight. This conclusion is supported by responses such as the following:

"As a dispatcher, I believe in the 'managing by exception' principle whereby I am shown issues that require my attention, whereas routine items are not dirrectly displayed to me, but are available for call up when I choose to do so."

"We have between 40-90 releases in a shift, so an alert would work best for our group."

Additional Information Needs

After answering the first question regarding form of access (alert vs. always present display), the dispatchers were asked about additional information that they might want to see incorporated into these screens. For Figure 2, the suggestions for possible types of information to add included:

- ATC preferred route
- Time and burn calculations for today's flight
- A traffic congestion index such as # of congested sectors transited
- Cost/time enroute and fuel requirements for alternative routings
- Out to off delays
- Reason for the route change
- Percentage of routes filed and flown successfully without ATC reroute

In addition to providing suggestions for including the information summarized above, more specific comments included:

"I think that screen has all the information I need for a quick and dirty risk assessment. … If I am working a DTW ATL flight, Figure 2 gives me an instantaneous assessment of which route I need to fuel for with a 99% probability (provided some other conditions are not a factor)."

"Out to off delays. … Seeing average out-off delays, assuming they are excessive, may make me look for other options (and why the out-off delays are consistently high)."

"I think you would need both planned speeds/altitudes along with altitude changes to filed routes/speeds/altitudes."

"Access to the previous few days' enroute weather."

The dispatchers were also asked whether they saw any problems with the content or design of specific screens.

Figure 2. Most of the dispatchers indicated they were happy with Figure 2, except for adding some of the additional information discussed earlier, making comments like

"In itself it doesn't contain much useful data. It should be incorporated into another screen."

"Simple and easy to read. It indicates your chance of being moved and the route to which you will be moved."

"Add the ability to break down by time. Sometimes flights are only consistently rerouted certain times of the day (i.e. due to crossing arrival/departure traffic at another airport)."

"It would be nice to show a mileage figure next to the route."

Figure 3. Most (but not all) of the dispatchers felt that Figure 3 required too much thought to use, emphasizing that they want to make decisions at a glance as much as possible:

"Too much info on charts. Difficult to understand in a short time. Dispatcher needs to see a trend or pattern, not raw data."

"I am not sure if we need EDCT's or Planned/Actual minutes when we have the difference. Rest is good."

"Historical table with EDCT, Dept. fix, holding, arrival fix info is confusing. The map with potentially overloaded sectors is good."

"This is great stuff. The information on this page would be very useful to me. I have no suggested changes."

"Good presentation. Table initially takes some study but once you get used to seeing it can focus in on important numbers."

Figure 4. The response was much more positive for most (but not all) dispatchers to the use of bar and scatter charts rather than tables to display data, but some dispatchers still had concerns about the specifics of these particular charts:

"Liked the charts. Easy/quick way to see trend. Also, good info and lots of info in small space. Desktop real estate is at a premium."

"This screen is more useful because I can gather much information at a quick glance. Bar graphs more useful than verbal statistical data."

"The screen is complex but easier to use than the tables. The bar graphs for off time delay and airborne delay indicate that regardless of what the flight planning computer says, your taxi out time and enroute time will not be as planned. We have one flight in particular where the actual enroute time exceeds the planned enroute time by 50% on an almost daily basis."

Figure 5. There was some disagreement about how much information to present about predicted sector loads. There was, however, general agreement that information on predicted sector loadings could be useful, assuming that it was reasonably accurate in the timeframe needed by dispatchers. There were also some suggestions for improving the details of this screen.

"Looks good. The ability to list flights in overloaded sectors would be useful."

"This is best screen for presenting all necessary info in a concise presentaion. No problems.

"I'm not sure that such detailed info regarding ATC sector in/out, duration, count, capacity is pertinent to a dispatcher. More basic info like 'will the sector I'm routing through be overloaded?' would be adequate."

"This table tells me that 3 ATC sectors are predicted to be saturated, but not how they will handle it. Will they delay or reroute UAL or USAIR and let my flight through or will I get the delay?"

"I have never been a big fan of predictive arrival information because of all the variables which impact a flight's actual arrival times at its destination. Whether it is a mechanical delay, flight attendants needing more ice, the boarding of additional meals, or whatever else may pop up, the statistical accuracy of any tool which predicts when a flight will get to its destination (or any other point along the route of flight) calculated before the actual gate departure is suspect. On the other hand, I can see where it could be useful to know something about the expected demand at an airport at the approximate time of my planned arrival. I would incorporate these estimates into my flight planning."

Likert Scale Questions

In addition to the open-ended questions summarized above, a number of Likert scale questions were asked. Figure 1 provides results for two very broad questions. In general, the more detailed questions were very consistent with the answers provided to the open-ended questions as summarized above.

Conclusions

The goal of these two studies was to gain insights into how predictions of air traffic activity as provided by FACET can be made useful and usable for airline dispatchers. Overall, there was a strong belief by the dispatchers studied that predictive data could be of substantial value to airline staff in making a variety of decisions, including decisions about:

- Fueling aircraft.
- Changing the routes or altitudes for flights, either preflight or while enroute.
- Expediting or delaying departure times.
- Negotiating with TFM to adjust traffic flows.
- Rebooking passengers.

The relevant time horizons require predictions ranging from 2 hours before departure to 10 minutes before departure to decisions made while a flight is enroute. Furthermore, while many of these decisions could be implemented effectively in the current NAS, new TFM procedures are likely to be needed to take full advantage of such predictive data.

The dispatchers emphasized the need not only to provide data regarding potential bottlenecks due to air traffic congestion, but also to provide insights into how these bottlenecks are likely to affect a particular flight. At present, the most effective way to accomplish this latter goal is to integrate predictive data with historical data.

In terms of usability, the dispatchers strongly emphasized the need to provide access to such information by exception (as alerts), and to provide it in a form that can usually be processed at a glance, but with additional details easily accessible for those cases where they are needed. A tool that requires substantial navigation through and analysis of the data is likely to be impractical for many of the tasks faced by dispatchers because of their high workload.

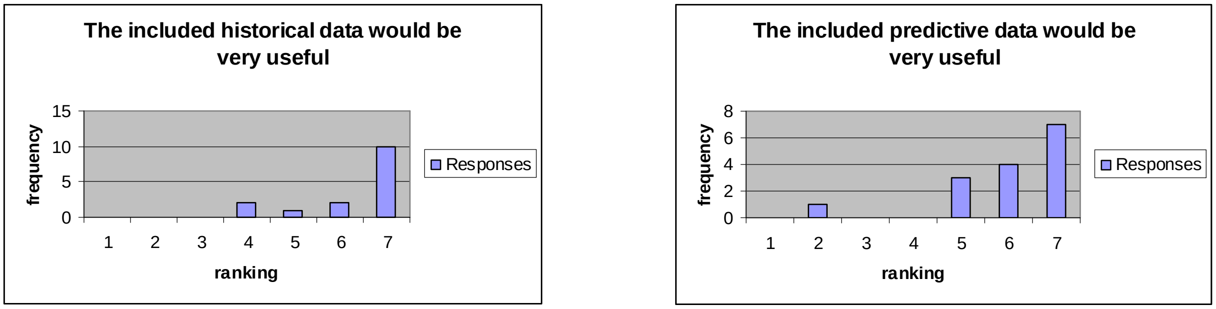


Figure 1. *Sample Likert Scale responses.*

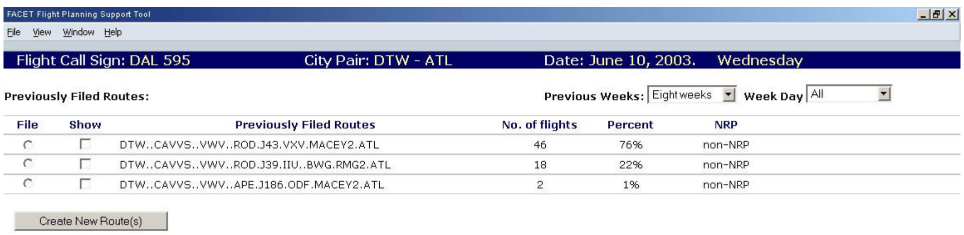


Figure 2. *Initial information on a specific flight.*

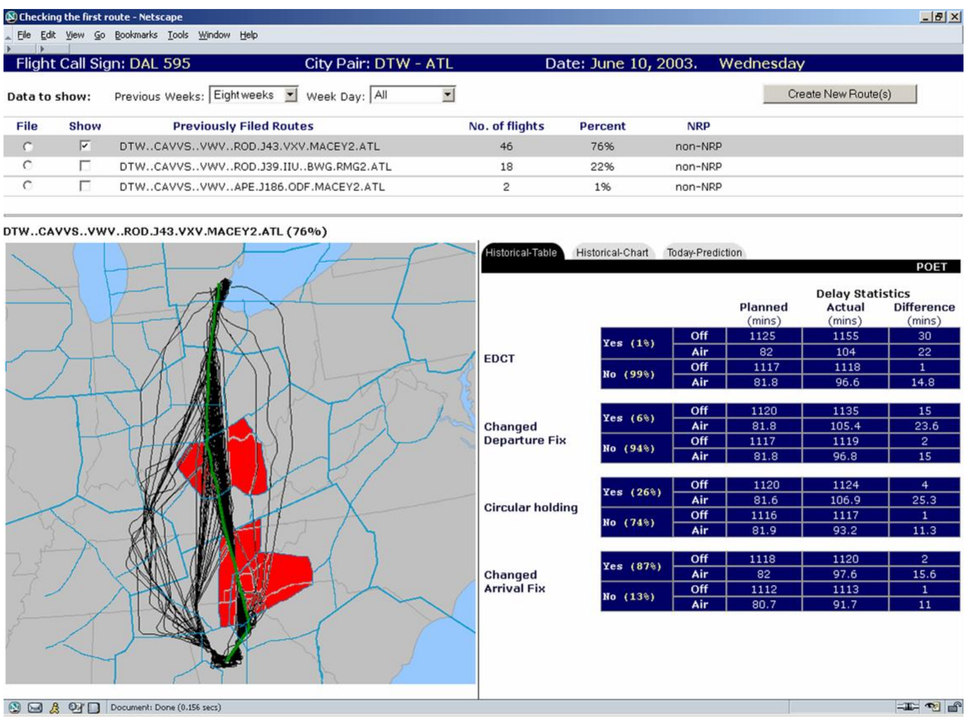


Figure 3. *Tabular presentation of historical flight performance data along with map display of FACET data and filed (green) vs. flown (black) routes. (The red sectors on the map represent sectors with "high" traffic volume.)*

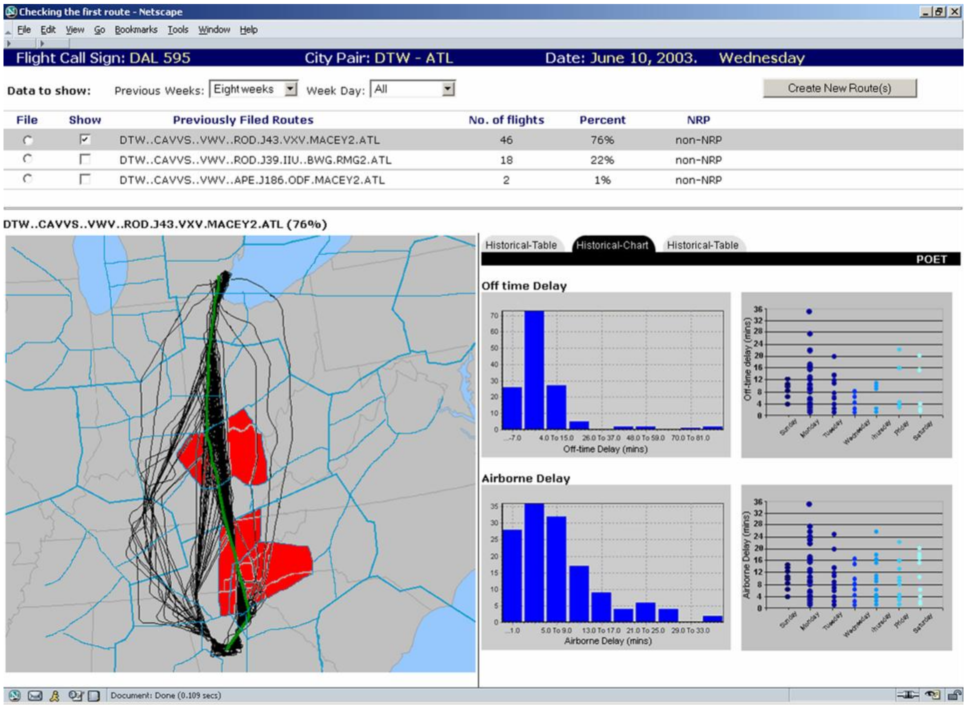


Figure 4. *Graphical presentation of historical flight performance data along with map display of FACET data and filed (green) vs. flown (black) routes. (The red sectors on the map represent sectors with "high" traffic volume.)*

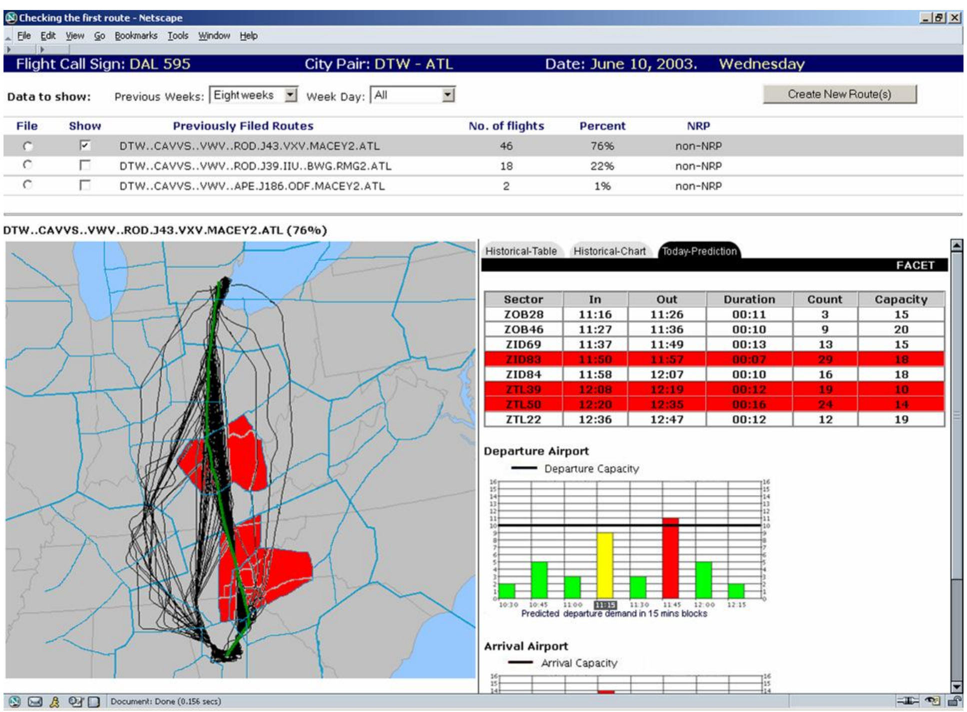


Figure 5. *Tabular presentation of FACET sector data, graphical display of FACET data on traffic demand at arrival and departure airports, and map display of FACET data and filed (green) vs. flown (black) routes. (The red sectors on the map represent sectors with "high" traffic volume.)*

A SYSTEMATIC APPROACH TO ADDRESSING HUMAN FACTORS CONSIDERATIONS IN THE DESIGN OF FLIGHT DECK COMPONENTS

Beth Lyall, Ph.D.
Greg Harron, M.S.
Research Integrations, Inc.
Tempe, AZ

The FAA Aircraft Certification Job Aid for Flight Deck Human Factors is a decision-support tool for addressing human factors considerations during the flight deck design portion of the aircraft certification process. The current version presents decision-support information related to the review of flight deck displays, controls, and systems. This tool provides a systematic approach for assessing human factors considerations related to the design of flight deck components.

Background

Most airplane certification projects for components of the airplane flight deck involve many decisions that are related to human factors principles. The FAA would like to have human factors considerations addressed as effectively and thoroughly as other considerations when making certification decisions. In addition, they would like to provide the resources for certification team members at all stages of the certification process to identify and possibly resolve any human factors issues that arise. There are Human Factors Specialists who are available to help with projects, but their workload precludes their availability for all day-to-day considerations within a project.

Challenge

There are many unique challenges that arise when reviewing designs for human factors considerations in a regulatory environment. One of the primary challenges is to know which human factors considerations should be addressed for any particular review or certification project. We have developed a decision-support tool for the FAA to meet these challenges. It is called the FAA Certification Job Aid for Flight Deck Human Factors (Job Aid).

Approach and Solution

The Job Aid is designed to provide quick and easy access to regulatory and human factors information that may be used by certification personnel for identifying and addressing human factors considerations for flight deck design. The current version of the Job Aid provides information addressing all human factors considerations related to the design review of displays, controls, and systems in the flight deck of large transport category aircraft. The set of human factors considerations provides a comprehensive way to address flight deck human factors in any certification project.

Approach to Job Aid Design

In developing the Job Aid, the certification personnel were interviewed and observed to determine when they would need human factors information and what their typical approach would be to acquire that information. It was determined that they needed human factors information when they were doing the following three tasks:

1. Reviewing related FAA regulations and guidance material,
2. Researching information related to a specific component (control, display, system, or equipment), and/or
3. Researching a specific human factors topic, such as clutter or the use of color coding.

The Job Aid has been structured to allow the certification team members to access information from any one of these three paths. When the user selects a particular regulatory or guidance document, component, or human factors topic; they will be provided with a list of related human factors considerations. This list of human factors considerations provides a systematic method of evaluating design and can serve as a general checklist during certification tasks.

Figure 1 depicts the design and structure of the Job Aid elements and their relation to each other.

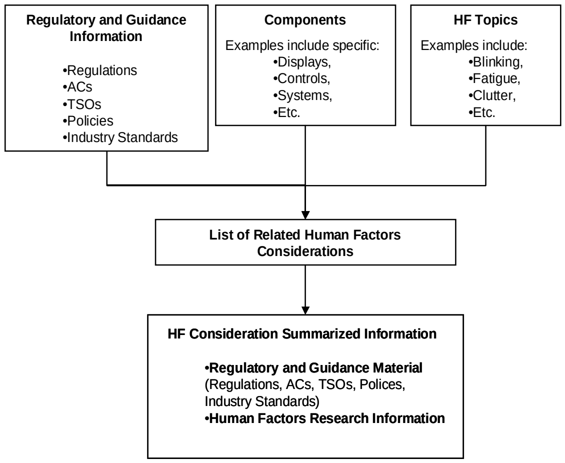


Figure 1. *Representation of the organizational structure of the Job Aid interface and information.*

Development of Human Factors Considerations

The biggest challenge in developing the set of human factors considerations was to define a set of considerations that could be used and understood by all the intended users of the Job Aid, and that would apply to current and future designs without the requirement of updating them as technology and innovation changes are made. The human factors considerations were developed by first reviewing the regulatory documents to identify the topics and organization of concepts related to human factors described. The terminology used in the human factors considerations was drawn from that used in the regulatory documents so that the concepts and descriptions included in the human factors considerations would be familiar to the Job Aid users.

Separate sets of human factors considerations have been developed related to display design, control design, and system design. The human factors considerations address the design issues of the component in isolation as well as design issues related to the integration of this component within the full flight deck environment. The Job Aid provides summaries of regulatory and guidance material as well as human factors research literature for each human factors consideration.

Examples of display-related human factors considerations are

- Information is visually accessible
- Information is understandable
- Timing of information presentation is appropriate
- Display appropriately attracts attention

Examples of control-related human factors considerations are

- Necessary controls are provided
- Control design prevents inadvertent operation
- Control function and method of operation are understandable
- Control is usable with related controls and displays

Examples of system-related human factors considerations are

- Pilot is provided necessary control over the system
- System operation or monitoring does not require excessive attention
- System logic and behavior are understandable
- System design minimizes potential for injury

After users are presented with a list of human factors considerations related to the document, component, or topic they have selected, the next step is to access the detailed information summaries that are provided for each human factors consideration.

Organization of Summarized Information

Due to the depth and breadth of the information provided for each human factors consideration, effective organization of the information is essential to making the information accessible to the users. For each human factors consideration, separate summary documents are included for a number of different types of documents. FAA regulatory and guidance documents have been summarized with separate summaries for Part 25 regulations, Advisory Circulars, Technical Standard Orders, Human Factors Policies, and Industry Standards. In addition, for each human factors consideration, a summary of non-regulatory research-based human factors information has been developed.

To help users further focus their search for information, the concepts described in the detailed human factors consideration summaries are organized into sections based on the type of information given. The summarized information is organized into the following sections:

- General Description,
- Design Strategies and Examples,
- Measures (design measures and pilot performance measures), and
- Trade-offs.

General Description. The General Description section includes a description of *all* design topics related to addressing the human factors consideration. The General Description section includes subtopic headings so that the user will be able to use this list as a comprehensive checklist for that human factors consideration. An example of a human factors consideration general topic and its subtopics is shown in Figure 2.

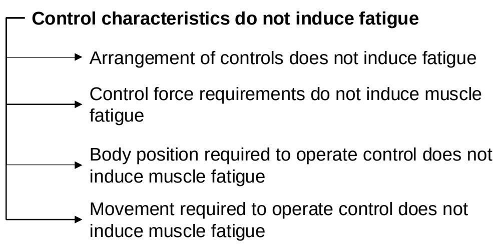


Figure 2. *Example of a human factors consideration topic and subtopics.*

Design Strategies and Examples. This section is also organized by the subtopics presented in the General Description section. It presents strategies that have been used to address each of the design subtopics. These strategies provide various options that may be used to address the particular subtopic of the human factors consideration along with their advantages and disadvantages. Specific examples of effective implementations of design strategies are also provided. Figure 3 shows an example of the design strategies for one of the subtopics presented in Figure 2.

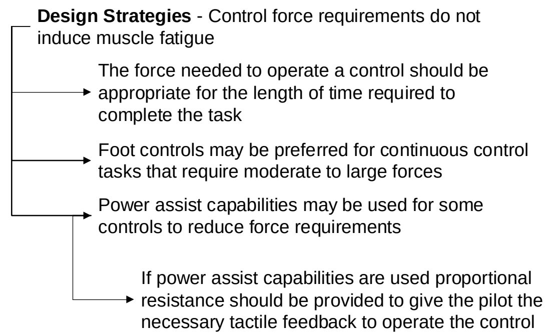


Figure 3. *Example of design strategies for a human factors consideration subtopic.*

Measures. This section provides information related to measuring the effectiveness of the design strategies implemented related to the human factors consideration topic or subtopics, as appropriate. The measures section consists of both design measures and pilot performance measures. Design measures refer to aspects of the design that may be measured to determine if they appropriately address the topic or subtopic. Examples of design measures might include character size measurements and display luminance values required to ensure adequate readability of display information.

The pilot performance measures section consists of aspects of pilot performance that can be measured to determine the effectiveness of the design related to the topic or subtopic. Examples of pilot performance measures include the pilot's speed and accuracy when completing tasks associated with using a flight deck component.

Figure 4 presents measures associated with the subtopic presented in Figures 3.

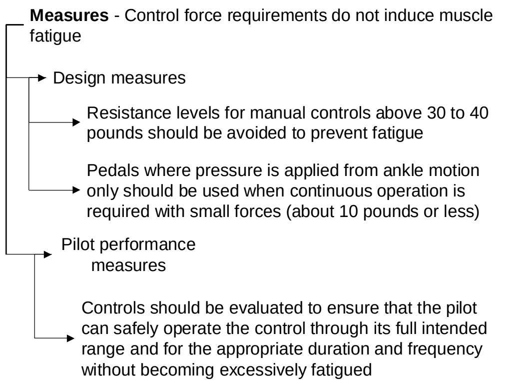


Figure 4. *Example of measures related to a human factors consideration subtopic.*

Trade-offs. This section provides descriptions of trade-offs that must be considered when making design decisions related to the topic or subtopics. The trade-off may be between two topics or subtopics in the same human factors consideration or it may be between a topic or subtopic in one human factors consideration and one related to another human factors consideration. Design decisions are never made in isolation and this section is meant to help the certification team members balance their decisions taking the whole design into consideration. Figure 5 presents the trade-offs associated with the human factors consideration topic presented in Figures 2, 3, and 4.

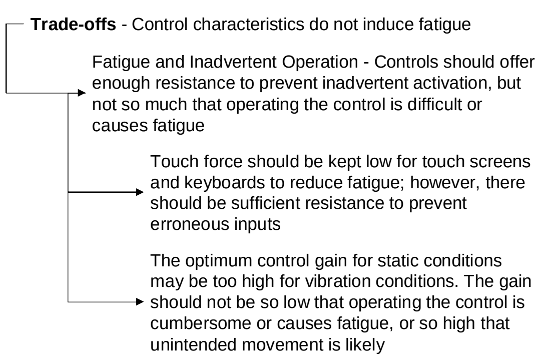


Figure 5. *Example of trade-offs for a human factors consideration topic.*

Implementation and Use of the Job Aid

As certification personnel systematically review the related human factors considerations, they are able to copy relevant text excerpts from the summaries and paste the information into a working document that can be used to compile relevant information from multiple human factors considerations and allow them to communicate issues and facilitate decision-making with the other certification team members and the applicant. The Job Aid does not provide direction on the certification decisions to be made, but it provides many related aspects of human factors information for the trained certification team members to use along with their other information to make their decisions.

The key to the effectiveness of the Job Aid is that it links the important aspects of human factors design review throughout the certification responsibility areas of the FAA. With this approach, the Job Aid can be used by all FAA certification team members to identify human factors considerations even if they only review small elements of a flight deck design.

Evaluation of the Job Aid as it is being used has shown that the human factors considerations are understandable and usable by engineers and other certification team members who have had little or no training in the science of human factors. The Job Aid has helped educate certification personnel who do not have human factors expertise and has allowed them to more effectively communicate with the Human Factors Specialists within the FAA. The Job Aid that is currently deployed in the FAA includes information for human factors considerations related to displays, controls, and systems. Future versions will include human factors considerations related to equipment, tasks and procedures, and testing assumptions. The current version focuses on certification projects for transport category aircraft; future versions will also include information related to small airplanes and rotorcraft.

Acknowledgements

This work was conducted under FAA Contract No. DTFAWA-04-C-00035. Our contract monitor is Dr. Tom McCloy and our technical sponsors are Dr. Kathy Abbott, Colleen Donovan, and Loran Haworth. The success of this work is greatly enhanced by the involvement and sound advice of each of them.

ANALYSIS METHODS FOR DETERMINING THE SAFETY CONSEQUENCES OF MIXED-FLEET FLYING

Beth Lyall, Ph.D.
Research Integrations, Inc.
Tempe, Arizona

Mike New, Ph.D.
Delta Air Lines
Atlanta, Georgia

This paper describes the development and implementation of analysis methods for identifying human factors safety vulnerabilities associated with the mixed-fleet flying of the Boeing 767-400 and 777. The results of the analysis were two sets of vulnerabilities: those with potentially critical pilot performance consequences that should be further tested, and those with minor consequences that should be considered when developing procedures and training for the mixed-fleet flying program. A longitudinal study was conducted that included data collection to address the potentially critical vulnerabilities. Examples of how vulnerabilities were addressed with video data from the study are presented and conclusions from the analysis are described.

Introduction

The current economic environment for airlines is causing them to consider all potential ways of reducing operating costs while maintaining high levels of safety. One of the ways that a few airlines have kept costs down is by flying only one type of airplane. This reduces training costs because pilots do not have to be trained as they transition within the airline to different airplanes. It also provides enhanced operational flexibility because any pilot can fly any airplane in the fleet. If there is a maintenance problem with one airplane, the pilots that were going to fly it will be able to fly any other airplane that the airline has available.

Airlines that have more than one type of airplane are looking into strategies for gaining some of the benefits of the single-type airlines. In hopes of realizing these benefits they are exploring is the possibility of mixed-fleet flying two or more airplane types. This means that the pilots would be simultaneously qualified in multiple airplane types and would be able to fly them indiscriminately as needed by the airline. Before such a program can be used by the airlines, however, the FAA must approve of their training and operations plan. Prior to FAA approval, the safety implications of the proposed mixed-fleet flying program are analyzed and a strategy is developed to mitigate any identified risks. Approval is given for a particular combination of airplanes and a specific training and operations program.

Airplane manufacturers are interested in mixed-fleet flying because it provides value to their customers if their airplanes can be flown by the same set of pilots. Traditionally, the manufacturer provides information to the FAA during certification of a new airplane model or type and the FAA gives approval for the type of training that must occur if that new airplane is flown with other models or types. At other times, the manufacturer may wish to explore the mixed-fleet flying option after an aircraft has been approved for operation. The current study follows this strategy. In this project, Boeing Airplane Company requested that the FAA assess the mixing of the Boeing 767-400 and 777 aircraft.

The Operator Differences Requirements Table includes details of differences between the designs of the airplanes and, for each difference, whether the difference results in flight characteristic changes or procedural changes, and what level of training and checking would be required at a minimum to address the difference.

The psychology of human behavior shows that it is not only the differences in airplane design that could cause difficulty, but problems may also result from similar designs that require different responses or procedures (Braune, 1989; Holding, 1987). The present study focuses primarily on this latter vulnerability as it relates to mixed-fleet flying. As part of this effort, we developed a systematic methodology for identifying and evaluating possible instances of negative transfer based on the mixed-fleet flying work by Lyall (1990). This approach has proven successful in its initial use and resulted in providing information about safety vulnerabilities, the potential for critical safety consequences, and possible mitigation strategies.

This paper describes the methodology; examples of specific safety vulnerabilities that were identified; and how the information can be used by airlines, airplane manufacturers, and the FAA.

Vulnerability Analysis Methodology

The objective of the vulnerability analysis was to identify pilot tasks for which there may be a potential for error as a result of mixing the airplanes. A full set of pilot tasks that was developed by the airline for their Advanced Qualification Program was used as the foundation of the analysis. The method was to document for each task the situations and actions required to accomplish the task on each airplane then analyze the similarities and differences between the airplanes, identifying potential vulnerabilities for pilot performance that could result from mixing the airplanes.

The documentation began by gathering information from the flight and training manuals for both airplanes. These manuals include normal and non-normal procedures and checklists. The information from the manuals was then expanded using interviews with pilots and instructors qualified on one or both of the airplanes and, if necessary, conducting directed sessions in simulators and training devices to understand how the pilots may need to operate differently in the two airplanes. This analysis resulted in a list of possible pilot performance vulnerabilities.

A "vulnerability" was identified if the required action or knowledge to accomplish a task was different in the two airplanes. Additionally, an item was categorized as a "major" vulnerability if it consisted of a critical behavior with safety consequences. All major vulnerabilities were further investigated as part of a separate analysis called the "longitudinal study."

Analysis Results

Possible vulnerabilities that needed to be further verified in the longitudinal study were identified for 19 tasks. Minor vulnerabilities that should be considered when updating training and procedures were identified for 17 tasks.

The most safety critical design difference that was discovered to have a vulnerability was the use of the Takeoff/Go Around (TOGA) switch for automatically advancing the thrust levers to takeoff power. The TOGA switch is used in three tasks: takeoff, missed approach, and rejected landing. The difference in design of the TOGA switch in the two airplanes may lead to a classic negative transfer situation because the tasks for which the switches are used are the same in the two airplanes, but the response the pilots must make in using the switches is different. In the 777 the TOGA switch is located only on the forward side of the thrust levers and would, therefore, be activated with the pilots fingers reaching forward and downward as the hand in on the thrust lever. The 777 TOGA switch is highlighted in the picture in Figure 1.

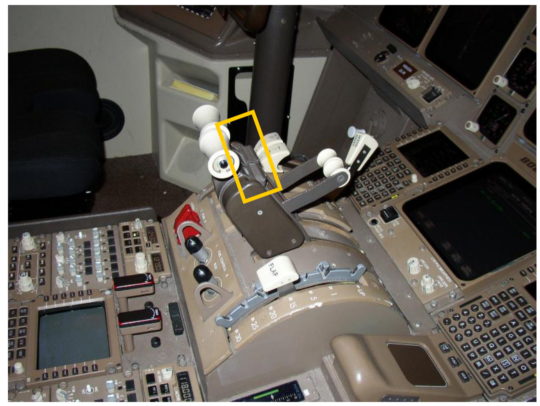


Figure 1. *Boeing 777 throttle quadrant with TOGA switch highlighted.*

On the 767-400, the TOGA switch can be activated from both the forward and aft positions related to the thrust levers. During our analysis we interviewed and observed several 767-400 pilots and instructors in the simulator and found that 100% of them used the aft TOGA switch exclusively. This gave us evidence that the difference in switch design could lead to safety consequences when the pilots who had habits formed in the 767-400 were flying the 777: They could take longer in activating the TOGA switch if they tried first to activate it at the aft of the thrust levers and had to realize that the switch was not there before activating the forward switch. We knew that the consequences of this response vulnerability had to be further investigated to determine its criticality and we recommended that it be included in determining the data to be collected in the longitudinal study.

Another example of a vulnerability that was further investigated in the longitudinal study is related to the engine start procedures. The 777 is designed so that both engines can be started simultaneously. It also has automated engine monitoring functions that alert the pilots if the engines are not starting properly. On the 767-400 the engines must be started separately, and the engine parameters must be monitored by the pilots to ensure a safe start. In our analysis process we found that the 777 pilots still closely monitored the engine status during start even though they had the assistance of the automation; however, they always quickly started both engines, which would not be appropriate in the 767-400. Therefore, we recommended monitoring the behavior of the 777 pilots when they started the 767-400 engines during the longitudinal study. In this case, the vulnerability was due to the differences in system logic and the resulting procedures used by the pilots and not to a difference in physical design characteristics. In fact, the controls used to start the engines were very similar, and this was noted as making it more likely that the pilots might perform the wrong behavior.

The minor vulnerabilities that were identified in the analysis were communicated to the airline to consider when developing training and procedures for their mixed-fleet flying program involving these two airplanes. Several of these vulnerabilities were due to the differences in the size of the two airplanes: The 777 is much larger in overall size and wing span than the 767-400. Because of this, there are several tasks during which the pilots must consider the size differences, and these differences are not necessarily evident while sitting in the flight deck. One example is during taxi procedures: The pilots in the 777 must ensure that they do not stop the airplane during a turn when the wheels of the main gear are still cocked in a turn. The main gear wheels of the 767-400 do not turn, so this is not a consideration on that airplane. Therefore, the vulnerability is for the pilots used to flying the 767-400 not monitoring the angle of the main gear wheels while taxiing the 777. This can be information given to the pilots during training. The consequences are wear and tear on the main gear assembly, but not safety related outcomes.

Verification of Vulnerabilities in Longitudinal Study

The scenarios and maneuvers used in the longitudinal study included elements to test all the safety critical vulnerabilities identified in the analysis. The study included pilot participants who were currently flying one of the airplane types and had never flown the other type. Participants were 10 captains and 10 first officers from each airplane, for a total of 40 pilots.

Longitudinal Study Design

The elements that make up this study are

- Simulator evaluation in current airplane
- Training in new airplane
- End-of-training simulator evaluation in new airplane
- 90-day simulator evaluation in new airplane
- 180-day simulator evaluation in new airplane

There were 10 crews per fleet (i.e., 10 crews were current in the Boeing 777 and received training in the Boeing 767-400 and 10 were current in the Boeing 767-400 and received training in the Boeing 777). Each crew consisted of a captain and a first officer. Half of the crews for each fleet were given a simulator evaluation in their current aircraft before any training began. Immediately after training, all of the crews were given a simulator evaluation in the newly trained aircraft. All the crews then returned to the line to fly their normal aircraft. At the 90-day interval, half of the crews from each fleet were given a simulator evaluation in the newly trained aircraft. At the 180-day interval, all the crews were given a simulator evaluation in the newly trained aircraft.

Pilot Training The training on the new airplane consisted of two systems training modules, a written systems evaluation, and five simulator training modules.

Evaluation Development The protocol and content for all evaluations were developed with input by the Boeing members of the MFF study team and FAA AEG inspectors. It was determined that the maneuvers to include in the evaluations should be those in the standard Appendix F required maneuvers list, with a few exceptions.

Evaluation Protocol The evaluation profile consisted of a line-oriented segment that included both crew members followed by a maneuvers evaluation segment that each of the participants flew separately. An instructor pilot served as the pilot-not-flying during the maneuvers segment to control for the possible confounding effects of having a non-qualified MFF participant serving that role. Because the instructors routinely serve this role during regular training and evaluation, they are familiar with the policy to neither help, nor hinder the pilot-flying during the evaluation.

Evaluation Data Collected The evaluation data listed in Table 1 was collected for each of the simulator evaluation segments. The data collected were the same for all evaluations.

The evaluation data were collected in the following ways:

- Instructor/evaluators provide yes/no assessments during the simulator evaluations. (Note: These assessments were also verified later during the video analysis.)
- Instructor/evaluators assign grades and identify errors during the simulator

evaluations (see Attachment 1 for details).
- Pilots complete modified NASA TLX worksheet assessment instruments following each module to measure perceived level of workload and provide workload ratings.
- Research Integrations conducts video analysis to collect data about the length of time required to complete the after landing checklist and the altitude at autopilot disengagement.

Video Analysis

All simulator sessions were video taped. The tapes were received and organized by evaluation modules, participant numbers, and participant positions (captain or first officer). The tapes were given unique numerical identities (1 – 113). The video analysis focused on the tasks for which there were safety critical vulnerabilities. We will describe here the analysis related to the use of the TOGA switch on the 777. Two tasks were analyzed: the missed approach at minimums and the rejected landing.

Hardware and software set up The video lab consisted of the following hardware:

- Apple Macintosh Powerbook with OS X
- 2 – Lacie 500 Gigabyte external hard drives
- 1 – Lacie 250 Gigabyte external hard drives
- 1 – Sony 20 inch color television
- 1 – LXI Video Cassette Recorder

The video lab used the following software to digitize and produce viewable files and DVDs

- Apple iMovie
- Apple iDVD
- Apple Quicktime

The LXI video cassette recorder's audio/video output was connected to the Powerbook audio/video input by a RCA cable consisting of left and right audio lines and a video line. All three Lacie External hard drives were daisy chained together using FireWire 800 cables and connected to the Powerbook through an iLink 6-to-4 pin cable.

Only the 777 tapes were digitized (45 in total), and an excel file was created capturing all of the data for each 777 tape. The tapes were imported and monitored to confirm:

- The proper functioning between VCR and iMovies as well as importation of video
- The tape's content corresponded accurately with the information on the tape spine and in the excel file:
  - Aircraft Type
  - Participant Position
- The video captured the entire simulation session without interruption or failure
- Any discrepancies were investigated and resolved

The file size for each digitized video tape was approximately 15 Gigabytes

Two segments of the simulation video were isolated for analysis: missed approach at minimums and rejected landing at approximately 100 feet

In capturing clips of the missed approach:

- All clips were edited to begin just before the airplane automated callout at "minimums."
- All clips were edited to end after the "gear up" callout by the pilot flying.

In capturing clips for the rejected landing

- Clips began just before the instructor command to "go around."
- Clips ended after the "gear up" call by the pilot flying.

In measuring the reaction times for the missed approach

- Reaction time began at the beginning "m" of the airplane automated "minimums" callout.
- Reaction time ended at the point of first observable hand or finger move going forward to activate the TOGA switch.

In measuring reaction times for the rejected landing

- Reaction time began with the instructor pronunciation of the "g" in "go around."
- Reaction time ended at the point of first observable hand or finger move going forward to activate the TOGA switch.

The file size for each of the clips was approximately 35 Megabytes.

The following information was recorded for each of the missed approach clips:

- Event time (duration of the clip)
- Reaction time for selecting TOGA
- Observations about how the pilot responded when selecting the TOGA switch: no hesitation, hesitation, went for aft button first, anticipated, etc.

The following information was recorded for each of the missed approach clips:

- Event time (duration of the clip)
- Reaction time for selecting TOGA

- Observations about how the pilot responded when selecting the TOGA switch
- Minimum altitude reached before climb out

The results of the video analysis show that there were several pilots who tried to select the aft TOGA switch when it was not there. The reaction times for these pilots were longer than for those who did not hesitate when choosing the forward TOGA switch. Figure 2 presents the numbers of pilots who reached for the aft TOGA switch before the forward switch while doing the missed approach. Figure 3 presents the same numbers for the rejected landing.

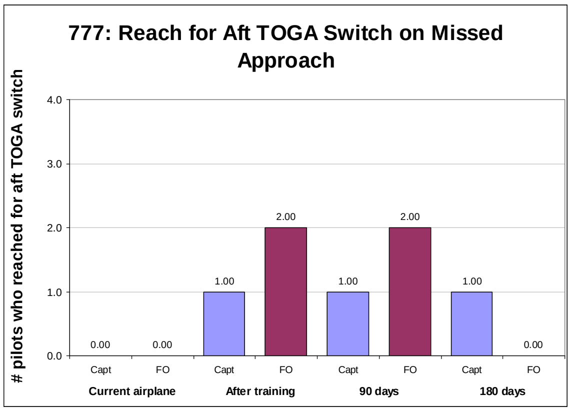


Figure 2. *Number of pilots who reached for the aft TOGA switch during the missed approach for each evaluation module.*

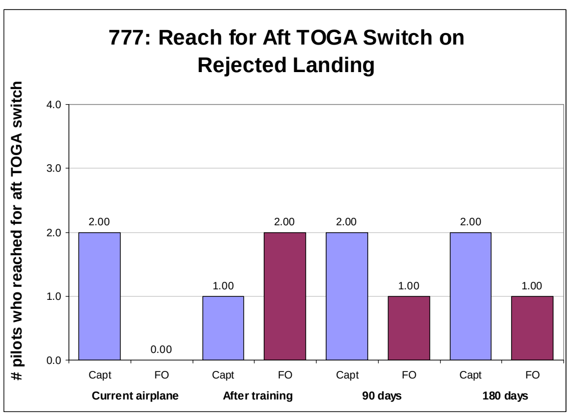


Figure 3. *Number of pilots who reached for the aft TOGA switch during the rejected landing for each evaluation module.*

Figure 4 shows the reaction times on the missed approach for those pilots who went for the aft switch first and those who went for the forward switch only. Figure 5 shows the same data for the rejected landing. The reaction times for the pilots who went for the aft switch first are significantly longer than those who directly activated the forward switch (both $p < .01$).

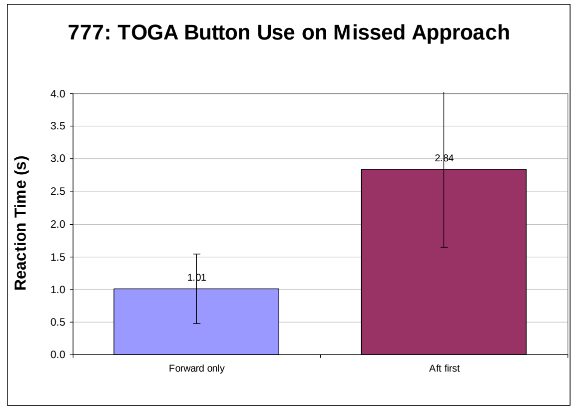


Figure 4. *Reaction times for activating the TOGA switch on the missed approach for pilots who reached for the forward switch only and the aft switch first.*

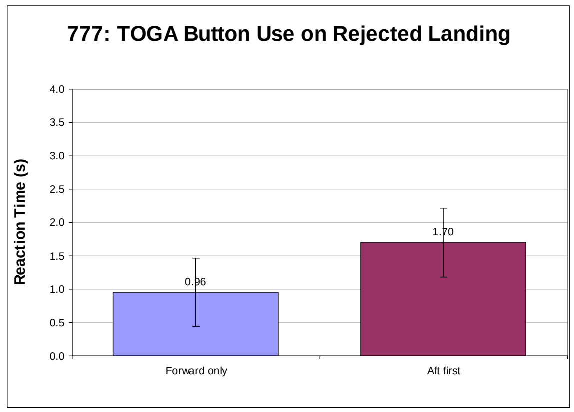


Figure 5. *Reaction times for activating the TOGA switch on the rejected landing for pilots who reached for the forward switch only and the aft switch first.*

These are the type of data that can be used to assess the criticality of vulnerabilities. It would be expected that negative transfer would occur more likely occur in situations that are unexpected and urgent or under time pressure. For the TOGA switch design vulnerability, the data show that more pilots, even those with current experience in the 777, reach for the aft TOGA switch in an unexpected situation like the rejected landing (Figure 3) than in a situation that can be anticipated like the missed approach at minimums (Figure 2).

It is also shown in Figures 4 and 5 that, as would be expected, the reaction times are longer when the pilot reaches for the aft switch first. The question for the FAA, manufacturers, and airline is whether the slower reaction time is enough to result in serious safety consequences. In this case, we included the missed approach and the rejected landing to be able to compare routine and time-critical reaction times.

Comparing Figures 4 and 5 shows that the pilots are quicker to respond in a time-critical situation like the rejected landing; however, the slower reaction times when reaching first for the aft TOGA switch make the safety ramifications of mixing these airplanes worth questioning whether there are mitigation strategies. For example, in this case a design change is one possible solution. The results of this analysis were passed on to the manufacturer and they are determining the requirements for adding the TOGA switch designed for the 767-400 to the 777.

We have shown through this study that a systematic vulnerability analysis can add significant value to the decision making required when determining the feasibility of mixed-fleet flying of any two airplane types or models. Because of the value added, the FAA and airplane manufacturer have requested that we be involved in future mixed-fleet flying assessments to perform the same type of analysis.

References

Braune, R.J. (1989). *The common/same type rating: Human factors and other issues* (Report No. 892229). Warrendale, PA: Society of Automotive Engineers.

Holding, D.H. (1987). Concepts of Training, in G. Salvendy (Ed.) *Handbook of Human Factors*. New York: Wiley.

Lyall, E.A. (1990). *The effects of mixed-fleet flying of the Boeing 737-200 and -300*. Doctoral thesis. Arizona State University.

Acknowledgements

This work was funded by FAA Contract No. DTFAWA-04-C-00035. Our contract monitor is Dr. Eleana Edens and our technical sponsor is Dr. Tom Longridge.

THE RELATIONSHIP BETWEEN CONTROLLER OPINIONS AND USE OF AN ATC DECISION SUPPORT TOOL

Carol A. Manning
Melanie Dennis
Federal Aviation Administration, CAMI
Oklahoma City, OK

The purpose of this study was to assess the relationship between controllers' opinions about the User Request Evaluation Tool (URET), an air traffic control decision support tool, and their use of various URET functions, including its flight data management capabilities, conflict prediction information, and problem-solving tools (e.g., trial planning). We expected that, compared with those who were less positive about URET, controllers who rated the tool more positively would use it more often in performing sector team duties. In 2002, formal observations were made of 181 en route controllers using URET at six facilities. URET display settings, usage, and use of automated flight strip equivalents were recorded. Controllers were also asked their opinions about the readability/usability of URET, changes in roles and communications between controller team members, their typical use of URET features, their perceptions of URET's effects on safety, workload, time required to perform tasks, and benefits provided to pilots. Dichotomously-coded answers to opinion questions were used as independent variables in t-tests. Dependent variables were counts of activities performed using paper, the Host computer, or URET. Controllers who performed more URET tasks thought the system required less time to use and were more positive about its effect on safety than those who performed fewer URET tasks. They were also more likely to indicate that they checked alerts and performed trial planning. Use of specific Aircraft List (ACL) functions was also related to controllers' likelihood of using URET's decision support capabilities. It is unclear whether controllers' familiarity with URET resulted in their positive opinions about the system or if having positive impressions made them want to use the system more frequently. It is also unclear whether increased use of the system would change controllers' opinions about URET. Regardless, these results indicate that there is a relationship between positive opinions and URET usage.

Introduction

We are interested in identifying factors that will predict the likelihood that air traffic controllers will use new automation tools to help them manage air traffic and perform their job duties more safely, effectively and efficiently. Two factors provide the reason for our interest. The first involves plans for new ATC automation tools to be introduced in the future. For example, Controller Pilot Data Link System (CPDLC), Automatic Dependent Surveillance-Broadcast (ADS-B), En Route Automation Modernization (ERAM), and other tools have been proposed. The second factor is the aging of the controller workforce resulting from the compressed hiring of controllers after the strike in 1981. We previously predicted that older controllers may have more trouble using new automation than younger controllers. However, Manning, Durso, Batsakes, Truitt, & Crutchfield (2003) found that en route controllers' age was not significantly related to their marking of paper flight strips, a process highly integrated in much of air traffic control. They also found that older controllers were able to use an alternative procedure for marking flight strips as easily as younger controllers. Moreover, Manning & Dennis (2004) found that age did not predict en route controllers' opinions about or use of an ATC decision support tool. Thus, while age has been found to be a factor in the use of some new technologies in the general population, it does not seem to be related to use of technology in ATC.

This study was conducted to identify other factors, besides age, that might predict automation usage. If age is not predictive of automation use, perhaps opinions about the automation might be. We assessed the relationship between controllers' opinions about the User Request Evaluation Tool (URET), an air traffic control decision support tool, and their use of certain URET functions. We expected that controllers who rated the tool more positively and, thus, considered it more useful, would use it more often than those who rated URET less positively.

Note that these data were collected soon after the introduction of the URET CCLD system. Several enhancements have been made to the system since that time as a result of controller feedback. Thus, comments provided at that time may no longer be relevant to the version of URET in the field today.

URET

URET is a decision support tool that provides the en route sector team with a conflict probe and electronic flight data management capabilities (FAA, 2001; FAA, 2005). A prototype version of URET was used on a daily basis at the Indianapolis and Memphis Air Route Traffic Control Centers (ARTCCs) for several years before the production system, called URET CCLD (Core Capability Limited Deployment) was introduced into 6 facilities in 2001 and 2002. URET is now operational at 10 en route facilities. URET provides timely and continuous detection of emerging problems, affording controllers with the opportunity to take action earlier and operate their sectors in a more strategic way. Aircraft-to-aircraft conflicts, for example, are detected up to 20 minutes in advance while aircraft-to-airspace problems are detected up to 40 minutes in the future. Controllers can use URET's trial planning capability to check a proposed flight plan amendment, such as a route change, for conflicts prior to issuing it as a clearance.

The primary URET display, the Aircraft List (ACL), consists of separate entries for aircraft currently under sector control as well as those predicted to enter the sector at some point within the next twenty minutes. ACL entries contain flight plan information, provide room for a controller to update information about issued clearances and show the status of URET-generated alerts for each aircraft. The ACL contains the same information and allows the controller to perform the same activities as the paper flight progress strips formerly used by en route controllers. Because it is difficult for controllers to take full advantage of the URET functionality while simultaneously managing paper flight strips, the requirement to use most flight strips is not in effect at en route sectors where URET is being used.

Method

In 2002, observations of 181 en route controllers were conducted at six facilities: the Kansas City, Chicago, Indianapolis, Memphis, Cleveland, and Washington ARTCCs. The controllers were observed while using URET. A checklist containing 79 items was available to guide the observation. Not every controller responded to all items.

Observers recorded URET display settings and controllers' use of URET automation functionality. Besides observing controllers' behavior, observers also asked them about their typical use of URET features, their opinions about the readability/ usability of URET ACL entries, their beliefs about the changes in roles and communications between controller team members, and their perceptions of URET's effects on safety and workload, the amount of time required to perform tasks, and the benefits they provide to pilots.

Answers to 4 opinion questions were recoded to be dichotomous (positive/negative) and were used as independent variables in t-tests. For t-tests, dependent variables were sums of counts of behaviors performed using URET. Independent variables were dichotomous responses to opinion questions. Chi-square analyses compared use of URET's decision support capabilities with dichotomous responses to opinion questions and observed use of specific URET functions.

Results

Analyses were conducted to investigate the relationships between three variables of interest: controller opinions about URET, use of URET to perform sector team tasks, and reported use of URET's decision support capabilities.

Opinions and URET usage

The first analysis looked at the relationship between controllers' opinions about URET and their observed use of the system. For this question, the independent variables were dichotomous codings (e.g., positive, negative) of controllers' opinions about URET's safety, workload required, time required to perform URET tasks, and benefits provided to pilots. The dependent variables were the number of flight strip equivalents or URET activities performed. URET activities were certain actions the controller could take using the system (see Table 1). These were looking or pointing at the ACL, acknowledging or coordinating route notifications, preferential routes, Unsuccessful Transmission Messages (UTMs), or Inappropriate Altitude for Direction of Flight (IAFDOF) indicators on the ACL, clicking to remove Ns from the bookkeeping box (when moving a new entry to the sorted list), deleting gray entries from the ACL (to remove aircraft that have been handed off to another sector), putting a checkmark in the bookkeeping box (to annotate an item to remember), highlighting ACL entries (to emphasize their importance), moving entries to the Special Attention Area, entering speeds or headings, opening or using the free text area, creating trial plans (for any reason), or using the Graphic Plan Display (GPD).

Table 1. *Specific URET activities recorded during observations*

Look at ACL
Point at ACL
Acknowledge or coordinate route notifications on ACL
Acknowledge or coordinate preferential routes on ACL
Acknowledge or coordinate Unsuccessful Transmission Messages (UTMs) on ACL
Acknowledge or coordinate Wrong Altitude for Direction of Flight indicators on ACL
Click to remove Ns from the bookkeeping box
Delete gray entries from the ACL
Put a checkmark in the bookkeeping box
Highlight ACL entries
Move entries to the Special Attention Area
Enter speeds or headings on the ACL
Open or use the free text area
Create trial plans (for any reason)
Use the Graphic Plan Display

Table 2 shows the relationship between controller opinions about URET and the sum of the different actions (described above) they were observed to perform while using URET. Significant differences were observed in the number of tasks controllers performed as a function of their opinions about the amount of time required to perform tasks using URET, the workload associated with their use of the system, URET's effect on safety, and additional services provided to pilots because of URET.

Controllers who thought URET saved time performed significantly more URET tasks than those who thought the system required the same amount or more time to use ($t(101)= 2.45$, $p < .02$). Controllers who made positive comments about URET's effect on safety performed more tasks using URET than those who expressed neutral or negative comments about its effects on safety ($t(140) = 2.46$, $p < .02$. No significant differences were observed in controllers' URET task performance as a function of their opinions about URET's effects on workload or whether they thought they provided additional services to pilots because of using URET.

Table 2. *Relationship between controller opinions about URET and number of different URET tasks performed*

Opinions about URET	N	Mean # tasks performed	SD	Sig
Amount of time required for use				*
Same or more	19	2.0	1.0	
Less	84	3.2	2.1	
Effects on workload				NS
Increased, no difference	31	2.7	2.0	
Reduced	107	3.2	2.1	
Effects on safety				*
Neutral, negative	73	2.8	2.0	
Positive	69	3.6	2.1	
Additional services to pilots?				NS
Yes	54	2.7	2.0	
No	39	2.1	1.6	

Opinions and reported use of URET's decision support capabilities

The second analysis looked at the relationship between controller opinions about URET and their reported use of URET's decision support capabilities. Table 3 shows the relationship between controllers' opinions about URET's safety, workload required, the time required for URET tasks, and benefits provided to pilots and their reported use of URET's decision support capabilities for checking alerts and performing trial planning.

Controllers who believed URET saved time were significantly more likely to report that they checked alerts than those who did not believe using URET saved time ($X^2(1) = 8.18$, $p < .01$). The reported use of trial planning as a function of the assessment of time saved by using URET was not significant. Controllers who believed that URET reduced their workload were more likely to check alerts ($X^2(1) = 6.58$, $p < .01$) and trial plan ($X^2(1) = 7.35$, $p < .01$) than controllers who thought using URET increased or did not change their workload. Controllers who made positive comments about URET's effect on safety were more likely to report checking alerts ($X^2(1) = 8.04$, $p < .01$) and trial planning ($X^2(1) = 13.00$, $p < .001$) than were controllers who made neutral or negative comments about the effect on safety. Controllers who said URET allowed them to provide additional services to pilots were significantly more likely to report checking alerts ($X^2(1) = 7.78$, $p < .01$) and using trial planning ($X^2(1) = 6.93$, $p < .01$) than those who did not believe they provided additional services to pilots when using URET.

Table 3. *Relationship between controller opinions about URET and reported use of its decision support capabilities*

	Use of URET's decision support capabilities					
	Check alerts?			Trial plan?		
Opinions about URET	#	%Y	%N	#	%Y	%N
Time required for use						
Same, more	19	37	63	19	42	58
Less	81	72	28	82	56	44
Effects on workload						
Increased, no difference	30	47	53	30	33	67
Reduced	103	72	28	101	61	39
Effects on safety						
Neutral, negative	72	56	44	69	39	61
Positive	65	79	22	64	70	30
Additional services to pilots?						
Yes	52	77	23	51	65	35
No	39	49	51	36	36	64

Number of task types performed using URET and reported use of its decision support capabilities

The third analysis examined the relationship between the number of different tasks performed using URET and controllers' reported use of its decision support capabilities. Table 4 shows the results. Controllers who reported that they at least sometimes checked alerts performed significantly more tasks than controllers who reported that they never checked alerts ($t(153)=2.3$, $p < .03$). Controllers who indicated that they at least occasionally used trial planning performed significantly more tasks using URET than those who indicated that they never used trial planning ($t(151) = 3.29$, $p < .01$).

Table 4. *Relationship between number of different URET tasks performed and reported use of its decision support capabilities*

Use of URET's decision support capabilities	N	Mean # tasks performed	SD	Sig
Check alerts?				*
Yes & sometimes	105	3.3	2.1	
No	50	2.5	2.0	
Use trial planning?				*
Yes & sometimes	85	3.6	2.2	
No	68	2.5	1.9	

Relationship of specific ACL functions to use of URET's decision support capabilities

The fourth analysis looked at the relationship between specific ACL functions used by controllers and their reported use of URET 's decision support capabilities. Table 5 shows that those who deleted gray entries from the ACL (a housekeeping task unrelated to planning) were significantly more likely to report that they checked alerts than those who did not delete gray entries ($X^2(1) = 12.39$, $p < .001$). However, deleting gray entries was not significantly related to reporting trial planning. Controllers who highlighted entries on the ACL were more likely to report that they used trial planning ($X^2(1) = 6.41$, $p < .02$) than those who did not. There was no significant relationship between highlighting entries and reporting that they checked alerts. Controllers who annotated speeds and headings were more likely to report that they checked alerts than those who did not ($X^2(1) = 5.56$, $p < .02$), but there was only a marginal relationship between annotation and trial planning ($X^2(1) = 3.08$, $p < .08$). Neither clicking the N to move an entry to the sorted list nor checking the bookkeeping box was related to reporting checking alerts or trial planning.

Table 5. *Relationship between controllers' use of specific ACL functions and reported use of URET's decision support capabilities*

	Use of URET's decision support capabilities					
	Check alerts?			Trial plan?		
Observed use of URET functions	#	%Y	%N	#	%Y	%N
Click N?						
Yes	57	75	25	57	63	37
No	89	62	38	87	48	52
Check box?						
Yes	38	71	29	37	62	38
No	101	67	33	98	54	46
Delete gray?						
Yes	116	72	28	116	54	46
No	17	29	71	15	40	60
Highlight?						
Yes	45	76	24	46	70	30
No	85	62	38	80	46	54
Annotate Speeds/Hdgs						
Yes	32	81	19	33	70	30
No	60	56	43	59	51	49

Discussion and Conclusions

Controllers who had more positive opinions about URET (e.g., saved time and enhanced safety) were observed performing more activities using the system. Positive opinions about URET were also related to whether controllers reported that they used the decision support functions of checking alerts generated by URET or trial planning the effects of making a proposed flight plan change. The number of different tasks performed using URET and the use of specific URET functions were related to whether controllers reported that they checked alerts or used trial planning. The results suggest that use of some URET functions associated with flight data management had a positive relationship with use of other functions related to its conflict probe capabilities.

Another result is that positive opinions about certain URET functions were related to increased usage of some aspects of URET. However, only some positive opinions about URET were predictive of URET usage. Moreover, positive opinions about URET did not predict use of all of URET's functions.

While attitudes have been found to be relevant to predicting behavior (Ajzen, 2001), other variables, as yet unidentified, may moderate the relationship. In this study, it is unclear if being more familiar with URET produced controllers' positive opinions about the system or if having positive impressions made them want to use the system more frequently. It is also unclear whether controllers' opinions about URET would improve after they increased their use of it. Regardless, these results indicate that there is some relationship between positive opinions about URET and controllers' use of the system. This suggests that it would be of value for developers of new ATC systems to assess controllers' opinions about system effectiveness and utility soon after implementation. The information may then be used to respond to concerns, either by providing clarifying information about how to use the system or by making modifications to it.

Acknowledgements

We would like to acknowledge Brenda Boone and Nikki Haase, Human Solutions, Inc., for their role in planning and executing the project. We would also like to acknowledge Skip Fisher and Dave Winokur, ATO-E, for reviewing and commenting on the manuscript.

References

Azjen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology*, *52*, 27–58.

Federal Aviation Administration. (2001). National Airspace System Capital Investment Plan: Fiscal Years 2002-2006. [Online]. Available: http://nasdocs.faa.gov/nasihtml/cip/cip02.pdf Washington, DC: Author.

Federal Aviation Administration. (2005). Free Flight Home Page: Tools: User Request Evaluation Tool (URET). [Online]. Available: http://ffp1.faa.gov/tools/tools_uret.asp.

Manning, C. A., & Dennis. M. (2004, May). Relationship between controller age and use of an ATC decision support tool. Paper presented at the 75th Annual Scientific Meeting of the Aerospace Medical Association, Anchorage, AL.

Manning, C. A., Durso, F. T., Batsakes, P., Truitt, T., & Crutchfield, J. (2003, April). Age, flight strip usage preferences, and strip marking. In R. S. Jensen (Ed.) Proceedings of the Twelfth International Symposium on Aviation Psychology. Dayton, OH: Wright State University, pp. 762-768.

TEAMS, TEAMWORK, AND AUTOMATION IN AIR TRAFFIC CONTROL

Philip W. Maynard and **Esa M. Rantanen**
Aviation Human Factors Division, University of Illinois at Urbana-Champaign
Savoy, IL, USA

Many recent initiatives involving social psychology applications in the aviation world have redoubled the interest in the concept of teams and teamwork. The importance of teamwork in airline cockpits, hailed as cockpit (or crew-) resource management (CRM), has been recognized for a relatively long time. It is also widely agreed that extensive and effective interaction among participants in the National Airspace System (NAS), pilots, air traffic controllers, and airline operations personnel, is tantamount to the daily successes of the nation's air transportation industry. Team aspects in air traffic control (ATC) are, however, much more convoluted than intra-cockpit teams or top-level teamwork between NAS elements. The ATC system involves a complicated network of facilities, technology, and personnel, which all must interact synergistically, often under time pressure, to ensure safe, efficient, and orderly flow of air traffic. It is perhaps due to this complexity that there has been a significant deficiency in research activity relating to teamwork in ATC. Yet, inadequate coordination between controllers has been considered a causal factor in a substantial proportion of low to moderate severity operational errors. Furthermore, automation tools developed for controllers are primarily focused on supporting the individual controller, while many, if not all of ATC functions are a team effort. In this paper we review the literature relevant to the team concept in the ATC domain, identify and characterize the different teams controllers belong to either simultaneously (e.g., intra- and inter-facility teams) or in different operational environments, and catalog the results from research literature as they pertain to the aforementioned teams in ATC and their specific characteristics. Our principal focus is on concepts such as taskload, workload, and situation awareness. Within this framework, we also map recent automation applications to ATC teams, hence highlighting their impact on the team dimension of human factors in ATC.

Introduction

It may be argued that the global air traffic control (ATC) system forms the largest singular team in aviation. It involves a complicated network of facilities, technology, and personnel which all must interact synergistically and often under severe time pressure to meet the ultimate objectives of ATC: safe, efficient, and orderly flow of air traffic from one location to another. Despite these inherent characteristics of the National Airspace System (NAS) in the U.S. and its international constituents, air traffic management (ATM) research with respect to automation-supported team decision-making has been fairly sparse. In fact, there has been a significant deficiency in objective scientific measures of ATC teamwork alone (Bailey & Thompson, 2000). Although this area is novel and still emerging, its further study in operational environments would potentially enhance the effective use of automation to aid team decision-making. Lapses in decision-making, coordination, and planning have been implicated in accidents and incidents alike and identified as latent problem areas in the NAS. According to a study by Rogers and Nye, coordination between controllers was considered a causal factor in 15% of low to moderate severity operational errors from 1988-1991 (Bailey, Broach, Thompson, Enos, 1999). Fortunately, a newfound interest has recently blossomed in this area due to the strong infusion of new technology into the ATC system and the foreseeable impact of automation on controllers' performance individually as well as on their interactions as members of various teams.

Implementation of automation in the worldwide ATC system's team of personnel to create safer and more efficient traffic management is easier said than done. There are many different members within the ATC system that have different and even conflicting strategic and tactical goals. Supporting these occasionally incongruent goals will require interfaces tailored to each position and job responsibility. Evaluating such a design has been described as a suitability assessment. A suitability assessment is the third part of a three-stage progressive assessment process geared towards systematic evaluation of system usability and task suitability of the system. Suitability assessments focus on the match between the system design and the user's task. A system is considered suitable if design features and functions support users well as they perform their tasks (Sanford et al., 1993). In this case, it is appropriate to evaluate a system in the context of the controllers' individual task of managing traffic while maintaining established team responsibilities.

However, there is a pervasive tradeoff between individual and team suitability assessments. Optimal automation for an individual is not always ideal for team performance (Hopkin, 1995). Tantamount in implementing automation as a 'team player' is a system that allows members of teams to maintain the best possible shared situation awareness (SA) and mental models. The importance of shared SA and mental models is forcefully explained by Wickens et al. (1998) and specifically in a study by Salas, Stout, and Cannon-Bowers (1994). This literature strongly asserts the need for shared mental models and SA as a linchpin for optimal team decision-making. In this paper, optimal team decision-making will be considered a function of a team's performance due to the interdependent nature of personnel and equipment in the NAS.

Current Automation Applications in ATC

We will discuss ATC team decision-making primarily in the context of the latest ATC automation tools: the Center-TRACON Automation System (CTAS) and its components. The User Request Evaluation Tool (URET) and Surface management System (SMS) are not part of the CTAS toolbox, but will also be discussed. More specifically, we will examine how these systems present information to the individual controller to support the underlying goals of the NAS and the more immediate objectives of the controller. CTAS is highly functional in that it features specific tools and interfaces for each control position. Such features, however, may conflict with established team norms and could undermine team performance (Hopkin, 1995). Furthermore, the automation that reduces team norms and standards will also disguise weakness or inconsistencies in team performance. This relates to actions of a controller troubleshooting being less visible to someone who might share the same problem. Thus, a significant aspect of implementing automation in the ATC domain would be evaluation of these consequences and their relationship to safety. These safety consequences currently are not directly apparent, however. Issues with automation in ATC include the extent to which team functions should be preserved and the importance of better identifying these functions so they aren't discovered to be necessary after the means to fulfill them have been automated out of the system (Hopkin, 1995). The tradeoffs of shared situational awareness with team and individual performance will be discussed for each component of the CTAS and their associated control positions.

CTAS

CTAS is a sophisticated system that consists of three major automation tools: the Traffic Management Advisor (TMA), Descent Advisor (DA), and the Final Approach Spacing Tool (FAST). On the Sheridan and Verplank (1978) scale, CTAS represents level 3 automation, where the controller is advised of action to take but has the option to disagree. In general, CTAS is primarily concerned with downstream flow and arrival traffic. As the name implies, it is utilized by both TRACON and en route center controllers. The TMA uses an interactive, menu driven timeline and a plan view display for Graphical User Interfaces (GUIs). The DA and FAST use graphical advisories and work in conjunction with the TMA kit.

SMS

Ground control of aircraft and scheduling of departure runways and times is handled by tower and ground controllers who are assisted by the SMS. The SMS advises and informs these controllers with runway balancing and departure schedule optimization (Walton, Quinn, & Atkins, 2002). The SMS features four types of displays at the controllers' disposal: maps, timelines, load graphs, and tables. The map display shows the location and direction of aircraft. The timeline predicts when an aircraft will be at a specific location (gate, runway etc), but does not show current aircraft position. Load graph displays show the current and forecast demand on airport resources. Meanwhile, the Flight and status tables provide flight-specific information (e.g., OUT and OFF times and departure runway; Atkins et al., 2004).

URET

The en route sector teams are assisted by the URET. This particular tool, which is independent of CTAS, allows these controllers to test scenarios of rerouting without having to mentally extrapolate the flight paths of numerous types of aircraft traveling at different speeds and altitudes. This tool takes into account, among other factors, aircraft performance and weather conditions to create a 4-dimensional flight profile of all aircraft. This is built into a human-computer interface, which is in textual and graphical format. The display provides the controller the ability to view aircraft routes and altitudes, predicted conflicts, and trial plan results. In addition, the point-and-click interface affords expedient entry and evaluation of trial plan route, altitude, or speed changes. Any changes made to the aircraft's flight plan are automatically updated in the central Host computer, which holds all flight plan information. (Walker et al., 2000). On a strategic and tactical level, this tool has the potential to reduce workload significantly and will potentially allow en route sector teams to effectively manage a larger taskload as traffic levels increase.

A Hypothetical Case Study

We will next discuss the implications of these automated tools on team aspects of ATC performance by a hypothetical case study, by following a generic flight from point A to point B through the NAS. The role of automation in coordination and collaboration between individual controllers will be highlighted.

Departure

After receiving their departure clearance from the clearance delivery controller, a commercial airline flight will typically be pushed back from the gate. At this point, the pilots' journey begins by talking to a ground controller, who will issue safe taxiing instructions to an active runway. At major airports, this is a complex job, as the intersections of taxiways and runways and sheer volume of traffic creates an intricate labyrinth of pavement and airplanes. Once the flight is at the intersection of a taxiway and the active runway, it is handed off to a tower controller, who issues the take-off clearance and is responsible for the initial departure sequencing. A human Traffic Management Coordinator (TMC), or supervisor, ensures smooth flow. Coordination demands concern appropriate assignment of runways to departing flights for least restricted climb-outs and to minimize the delays between successive departures, which are necessary for safe separation and wake turbulence avoidance.

Control of aircraft on the ground at the airport is augmented by the SMS. A simulation study by Walton, et. al (2002) has revealed several points to note in how the controllers who control departures utilize this system to make decisions and handle aircraft. This particular automation system has potentially negative effects on shared situational awareness, however. The root cause for this could be a result of different goals and displays that support these goals. The TMC has strategic goals, while the ground and local controllers have predominantly tactical goals (Walton et. al, 2002). Therefore, ground and tower controllers will allocate their attention to the information available to them in order to suit their tactical goals, while the TMC is looking at the big picture and a different display to make strategic decisions. The consequence is a decrease in shared situation awareness. Walton et al., (2002) reported controllers experiencing information overload and over-redundancy, which may cause cues to become selectively filtered and processed according to salience (Wickens & Hollands, 2000). Also, when workload increases, controllers will further channel their actions and attention to support their job responsibilities (Hopkin, 1995), thus decreasing teamwork. This teamwork detriment could result in action decisions being made with suboptimal SA. More specifically, the controllers' ability to perceive a change in the environment, understand it, and predict the future state will be compromised. In such an unforgiving field as air traffic management, making decisions based on suboptimal SA carries potentially dangerous consequences.

To compound the situation, the simulator study revealed reliability problems in the advisories, which were partially due to algorithm problems (Walton et al., 2002). Essentially, the same information with different meaning to certain personnel is going to have implications for how controllers in charge of departure flow interact with those who actually manipulate the airplanes to create that flow. This situation is further complicated by less than acceptable user ratings of automation reliability.

Enroute

Once the aircraft is airborne, it is handed off to the departure controller, who will place the airplane on a departure procedure to route them out of the terminal airspace and into the en route structure of the NAS. Next, the aircraft will be handed off to a controller in an air route traffic control center (ARTCC, or center). Most of the flight will be spent interacting with a series of center controllers who control a 3-dimensional block of airspace known as a sector.

Typically, sectors in ARTCCs are controlled by a team of two controllers, a radar controller, or an "R-side" controller, while and a flight data controller, or "D-side" controller. The R-side controller is typically charged with maintaining separation of the airplanes in the sector, and this controller is the one who communicates verbally with aircraft over the radio. The D-side controller is responsible for coordinating the transfer of control of aircraft to other sectors or facilities, as well as providing a second opinion and safety mechanism for the R-side controller (Bailey & Willems, 2002).

The primary automation tool available for en route sector teams, URET, fosters solid team decision-making within the team. The R-side controller receives a re-routing request from an aircraft and gives the information to the D-side controller, who has access to the URET. After testing the scenario or creating a more acceptable one, the D-side controller will inform his or her counterpart of the situation (Wickens et al, 1998). This system fosters a shared mental model because the D-side controller can only work with the information he/she is given by the R

Side controller, which is a manifestation of their understanding of the situation.

One potential issue for further investigation in this case is the compatibility of the URET with the CTAS's Traffic Management Advisor (TMA). This tool augments the enroute and TRACON traffic management controllers. The TMA develops a plan for each individual aircraft and sequences multiple aircraft arrivals in relationship with each other (Wickens et al, 1998). If the en route sector teams are routing aircraft in a manner contrary to the TMA or supervisors' plan, the TMA when combined with the URET could reallocate workload for the users. On different levels, this idea has been expressed consistently in the literature (Wickens et al, 1998, Sanford et al, 1999, Sanford et al, 1993). Essentially, the operators of the URET and TMA would have to effectively communicate to ensure their goals and SA is consistent.

The last sector to handle an aircraft before it re-enters the terminal environment works with the Descent Advisor (DA) CTAS tool. This tool assists controllers by structuring advisories to create a seamless transition from the en route phase of flight to the arrival. There is often a bottleneck at this point in the system, and this automation is an effort to mitigate the arrival bottleneck.

The DA advisories include fuel-efficient top-of-descent (TOD) points, speed profiles, altitudes, and vectors. Conflict resolution and management conformance advisories are supported automatically or semi-automatically through scenario planning (Sanford et al, 1999). This system contributes to team decision-making in that its primary objective is to integrate the notoriously separate tasks of en route control to arrival sequencing. Also, it adds a third dimension by ensuring the traffic management personnel's policy is being implemented in the system's output. In theory, the system integrates all involved parties' goals to create common solutions.

Despite the commendable goals of the DA tool, it is not free of automation-related human factors concerns. A primary concern in the evaluation literature is that of redistributing workload. The DA essentially reallocates the human's role in the system by forcing them to perform primarily strategic control action as opposed to the tactical control they exercised previously (Sanford et al, 1999). Further problems with reallocating workload lie in the new tasks required of the human operator. In the case of the DA, the system performs all tactical decisions in the form of a level 3 automation advisory. As the human's role shifts to performing strategic tasks, their mental model is now sub-optimal due to the inherent fact that automation represents data in terms of a direct visualization (Pea, 1993; Salomon, 1993; Wickens, 1992). The operators are less informed about how the tactical decisions were made, thus their mental model is degraded because they are not actually thinking about the situation. Because effective strategic decision-making is comprised of numerous tactical decisions, fully optimal strategic decision-making may possibly be hindered by the DA. An automated system such as the DA is consistent with the assertion that positive automation attributes of low workload and good prediction have implications for a good mental model (Wickens, 1992).

Arrival

As the airplane transitions from the en route phase of flight to the arrival, the aircraft control is handed over to the TRACON controller. These controllers primarily sequence aircraft for approaches and issue landing clearances. Once the airplane is clear of the active runway on a taxiway, the control is once again passed to the Ground controller for the taxi clearance to the gate.

Efficient aircraft arrival is aided by the Final Approach Spacing Tool (FAST) component of CTAS. Currently, a passive level 3 automation version of FAST is used, referred to as pFAST. The pFast utilizes advanced logic and algorithms to sequence aircraft by advising the controller. It also performs runway allocation tasks. This tool aids a task that is notorious for very high workload and even has made a significant impact on improving throughput. Dallas-Fort Worth reported a 9-13 % increase in throughput as a result of pFAST implementation (Quinn & Robinson, 2000). Proper flow management by supervisors should enhance the effectiveness of pFAST.

The interdependent nature of air traffic control teams and automation's effect on their performance is quite evident. From a review of automated systems in ATC, there are many positive attributes associated with their operational implementation. There are also positive aspects in the design and evaluations of these systems, as experienced air traffic controllers are highly involved in the design and evaluation processes (Quinn & Robinson, 2000; Harwood, 1993; Sanford et al., 1993,1999; Walton et al., 2002). The overall system benefits are not without cost to shared SA at some point in the system. This shared SA is the linchpin in effective and optimal decision-making with automated traffic management tools.

Discussion

The current automation interventions to aid the air traffic controller have caused the ATC teamwork concept to evolve and adapt. Although each system is lauded for key positive technological and task-driven features, several factors indicate a guarded approach to automation implementation should be followed.

The positive aspects of innovative automation approaches to ATC are the benefits for documented throughput and alleviation of some time-pressured, high-workload problem solving tasks which controllers would find increasingly difficult in the ever-growing volume of future air traffic. Some automation, such as the URET, actually does foster strong team decision-making on account of both individuals utilizing the same information to compose a mental model and conduct action decisions with the same information at their disposal.

However, the drawbacks to some automated approaches suggest more work will need to be accomplished in determining optimal automation suitability for the individual and the team's task environment. This involves a strong foundation of understanding more precisely how air traffic control teams interact and how automation can best support this interaction. More specifically, differences in strategic and tactical goals between two different tiers in the ATC system can invite difficulty in sharing SA. Also, this type of situation could result from the interaction of two independently designed automation systems. Other issues include the redistribution of workload among the various ATC specialists and managers. Furthermore, automation supporting direct visualization can hinder a controller's problem-solving skills (Pea, 1993; Salomon, 1993; Wickens, 1992). On the technological side, unreliable automation can produce an entirely separate set of human performance problems, particularly involving trust and reliability issues.

Despite the drawbacks, automation and advanced technology implementation has much potential to assist controllers and improve safety. However, this safety improvement potential can be realized by constructively analyzing the strengths and weaknesses of automation in the context of the concept of ATC as a single large team.

Conclusion

Areas for future research include the cost-benefit analyses of automated tools and the specific effect of their implementation on controller workload and team decision-making. It is a realistic possibility that future air traffic demands will dictate that controllers and traffic managers must operate with a certain amount of a decrease in shared situational awareness in order to meet the demands. These studies performed in the context of future air traffic demands will be beneficial in ensuring excessively severe latent issues will not manifest themselves later. Given the complexity of the ATC system, these evaluations are difficult but necessary in preparing for the future air traffic demands.

The task of managing traffic in the NAS is certainly a daunting one, as air traffic is projected to continue to increase in the near future. Effective understanding of how ATC teams function and how to best support team coordination with team-centered automation approaches will improve shared situational awareness of all team members and will consequently enhance the safety and efficiency of operations.

References

Atkins, S., Y. Jung, C. Brinton, L. Stell, T. Carniol, & Rogowski, S. (2004) Sufrace Management Field Trial Results. *AIAA 4th Aviation Technology, Integration, and Operations (ATIO) Forum*, Chicago, IL, September 20-22.

Bailey, L.L., Broach, D.M., Thompson, R.C., & Enos, R.J. (1999). *Controller teamwork evaluation and assessment methodology: A scenario calibration study* (DOT/FAA/AM-99/24). Washington, DC: Federal Aviation Administration.

Bailey, L. L., & Thompson, R.C. (2000). *The effects of performance feedback on air traffic control team coordination* (DOT/FAA/AM-00/25). Washington, DC: Federal Aviation Administration

Bailey, L. L., & Willems, B.F. (2002). *The moderator effects of taskload on the interplay between en route intra-sector team communications, situation awareness, and mental workload* (DOT/FAA/AM-02/18). Washington, DC: Federal Aviation Administration.

Harwood, K., (1993). Defining human-centered system issues for verifying and validating air traffic control systems. In J. Wise, Hopkin, V. D., Stager, P. (Eds.), *Verification and validation of complex integrated human machine systems*. Berlin: Springer-Verlag.

Hopkin, V.D. (1995) *Human factors in air traffic control*. London: Taylor & Francis.

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 47-87). Cambridge, UK: Cambridge University Press.

Quinn, C. M., & Robinson, J. E., III, (2000). A human factors evaluation of active final approach spacing tool concepts. *3rd USA/Europe Air Traffic Management R&D Seminar*. Napoli, Italy, 13-16 June.

Rodger, M. & Nye, L. (1994). Factors associated with the severity of operational errors at ARTCC. In M. Rodgers (Ed.), *An examination of the operational database for ARTCC* (pp. 11-15). (DOT/FAA/AM-93/22). Washington, DC: Department of Transportation/Federal Aviation Administration.

Salas, E., Stout, R., & Cannon-Bowers, J. (1994). The role of shared mental models in developing shared situational awareness. In R. Gilson, D. Garland, & J. Koonce (Eds.), *Situational awareness in complex systems* (pp. 297-304). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Sanford, B.D., Harwood, K., Nowlin, S., Bergeron, H., Heinrichs, H., Wells, G., Hart, M. (1993). Center/TRACON automation system: Development and evaluation in the field. *Proc. 38th Annual Air Traffic Control Association Conference* (pp. 238-245). Washington, DC: ATCA.

Sanford, B. D., Smith, N., Lee, K. K., & Green, S. M. (1999). Decision-aiding automation for the en-route controller: A human factors field evaluation. *Presented at the Tenth International Symposium on Aviation Psychology*, April 1999.

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. Cambridge, MA: MIT Man-Machine Systems laboratory Report

Salomon, G. (1993). No distribution without individuals' cognition: a dynamic interactional view. In G. Salomon (Ed.), *Distributed Cognitions*: *Psychological and Educational Considerations*, (pp. 111-138). Cambridge, UK: Cambridge University Press.

Walker, M.G., Long, P., & Lowry, N., (2000) *URET daily use metrics and benefits analysis–Progress report* (MP00W0000164). McLean, VA: MITRE Corporation.

Walton, D., Quinn, C. & Atkins, S. (2002). Human factors lessons learned from a surface management system simulation. *AIAA Aircraft Technology, Integration and Operations (ATIO) Conference*, Los Angeles, CA, October 1-3.

Wickens, C.D., & Hollands, J.G. (2000). *Engineering Psychology and Human Performance*. Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D. (1992). Virtual reality and education. *Proceeding of the 1992 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 842-847). IEEE.

Wickens, C.D., Mavor, M., Parasuraman, R., & McGee, J., *The Future of Air Traffic Control*. Washington, DC: National Academy of Sciences

EVALUATION OF PILOT AND RUNWAY CHARACTERISTICS ASSOCIATED WITH RUNWAY INCURSIONS

Kathleen McGarry
Titan Corporation
Washington, DC

Gregory Niehus
Titan Corporation
Mays Landing, NJ

A study was conducted to determine possible pilot and airport characteristics that could be used to predict surface incidents (SI) committed by pilots. The study was conducted by analyzing the videotape data from a previous simulation study (Surface Technology Assessment Product Team, 2004) that tested the ability of dynamic taxiway message signs called Addressable Message Boards (AMB) to enhance pilot situation awareness (SA) and reduce the likelihood of SIs at controlled airports. The current study did not take into consideration the impact of AMBs on SIs, but specifically focused on pilot and scenario characteristics. The results of the study indicate that pilots who committed SIs had logged fewer hours in the past six months than the pilots that did not commit a SI. In addition, pilots who committed SIs also had less experience at controlled airports than pilots who did not commit a SI. Pilots committing SIs also had lower situation awareness and higher levels of workload. It is likely that the combination of less recent flight experience and less experience at controlled airports were the cause of increased workload and lower SA for some pilots. The resultant increased workload and decreased SA led to a higher likelihood of these pilots committing surface incidents.

Introduction

Between FY 2000 and 2003, the National Airspace System (NAS) managed approximately 262 million flight operations. Of these, 1,475 resulted in runway incursions. That averages out to about five runway incursions per million operations (Office of Runway Safety, 2004). The FAA is evaluating and identifying strategies and emerging technologies for increasing runway safety.

The FAA defines a runway incursion as "any occurrence at an airport involving an aircraft, vehicle, person or object on the ground that creates a collision hazard or results in a loss of separation with an aircraft taking off, intending to take off, landing, or intending to land." The FAA categorizes runway incursions into three error types: operational errors, pilot deviations, and vehicle/pedestrian deviations (Office of Runway Safety, 2004). An operational error is defined as an action of an air traffic controller that results in an aircraft landing or departing on a closed runway, or less than the required separation distance between two aircraft, or between an aircraft and another obstacle (such as a vehicle, equipment or personnel). A pilot deviation is an action taken by a pilot that violates any Federal Aviation Regulation, such as if a pilot fails to follow air traffic control (ATC) instructions to hold short and not cross an active runway. A vehicle/pedestrian deviation occurs when pedestrians, vehicles or other objects interfere with airport operations by entering or moving on the runway movement area without authorization (Office of Runway Safety).

The FAA has developed five operational dimensions that affect runway incursions: available reaction time; evasive or corrective action; environmental conditions; speed of aircraft and/or vehicle and; proximity of aircraft and/or vehicle (Office of Runway Safety, 2001). These five dimensions were involved with the development of runway incursion categories based on severity. These categories are:

- Accident: A runway incursion that resulted in a collision.
- A: Separation decreases and participants take extreme action to narrowly avoid a collision.
- B: Separation decreases, and there is significant potential for collision.
- C: Separation decreases, but there is enough time and distance to avoid a collision.
- D: Little or no chance of a collision, but meets the definition of incursion

One of the main safety goals of the FAA is to reduce the rate of runway incursions. In addition to runway incursions, there are also SIs. An SI is any event "where unauthorized or unapproved movement occurs within the movement area associated with the operation of an aircraft that affects or could affect the safety of flight"

(Air Traffic Evaluations and Investigations Staff, 2002). SIs result from the same things as runway incursions: pilot deviations, operational errors, vehicle or pedestrian deviations and operational deviations. These surface incidents would result in a runway incursion if there was another vehicle in conflict at the time the incident occurs (Koenig, 1995).

With the increasing demand on the NAS for air travel, airport surfaces are becoming more crowded. While many air carriers suffered decreases in air traffic following September 11, 2001, demand for air travel is on the rise again. Congestion at airports is a major safety concern, and finding ways to prevent runway incursions is an area of research interest. Runway safety is managed by the pilots and air traffic controllers, who use visual and radio communications to maintain separation on the airport movement area. The airport movement area is the area where aircraft and vehicles are required to have permission from the air traffic control tower to operate (Pope, 1990). The path they are to follow is given to them by radio communication from the ground controller. The ground controller must maintain an awareness of where all aircraft are that they have given taxi instructions to. This is to avoid giving an aircraft a route to follow that will put them in the path of another aircraft or vehicle. The pilot and flight crew must either write down the taxi instructions or memorize them, and then follow the airport signs to their destination (Young and Jones, 1998). Often, position awareness is determined by both pilots and controllers through visual scans of the airport surface, using signs, lights, and pavement markings. Often, pilots use a paper surface map to assist them in determining position awareness. This may be especially true at unfamiliar airports (Young, et al, 1998).

The fact that much of positional awareness is based on visual scans of the airport surface makes it difficult to maintain awareness if visibility drops, if there is uncertainty regarding the correct path, or if there are obstacles such as other traffic in the way. This is true especially at unfamiliar airports. Position uncertainty can cause pilots to slow down until they gain a better idea of their position. It can also cause them to continue at speed, but with a lowered level of comfort (Young, et al 1998). The way that route information is given - voice communications - can be unsafe, if the communications are misunderstood or unclear. Pilots occasionally have difficulty understanding clearances, especially if the airport has a complex configuration. The pilots may also mis-hear messages intended for another aircraft, especially if the call sign for the intended aircraft is similar. Pilots can ‘hear’ a clearance that is expected, even if it is not given. They may act on their expectations, and not on the actual clearance given (Pope, 1990).

There are many factors that go into the cause of a runway incursion. The factors that go into the human error that cause runway incursions have been examined in previous research. The factors include: how pilots navigate the airport surface; how the runways and taxiways are identified (signs, lights, etc); communications (message content and message delivery); pilot and controller memory; situation awareness; lack of standardization; variability of training; pilots knowing where they are located; pilots knowing where other traffic is located; pilots knowing where to go on the airport surface (Jones, 2002; Adam, Lentz & Blair, 1992).

Study Objectives

The purpose of this study was to examine questions emanating from the simulation data collected during the AMB study involving runway incursion prevention technology (Surface Technology Assessment Product Team, 2004). The current research endeavored to identify factors that can be used to predict and prevent runway incursions based on pilot performance in the AMB taxi scenarios. A second objective was to examine whether the methodology utilized in the previous study can be used to learn more about SIs in the NAS to predict airport surface safety risks. The study investigated whether or not the methodology used in the AMB study would be useful in attempting to predict runway incursions based on knowledge of pilots and the scenarios that they typically experience.

Also of interest was whether or not violations of hold short instructions were predictable from the AMB scenarios. If they are, ‘typical’ surface scenarios for an airport could be assessed, and used in conjunction with knowledge about the pilot population to predict ‘typical’ airport safety risk areas. The scenario characteristics examined included: surface traffic, airport layout, unexpected surface characteristics, and radio communication.

Method

This research was an extension of a previous simulation study examining the use of dynamic message signs as a method of mitigating runway incursions (Surface Technology Assessment Product Team, 2004). This study extracted data from the AMB videotapes for use in the analysis of pilot characteristics and performance to look for possible causal factors and predictors of runway incursions and SIs.

Researchers used video and questionnaire data from the 28 pilots who participated in the AMB study. The pilots performed taxi operations in a Cessna 421 in a simulated environment in six scenarios from four different airports. The airports, selected based on the Runway Incursion Assessment Report (FAA, TAT 2002), were: Long Beach Airport, California (LGB); Crystal Airport, Minnesota (MIC); Flying Cloud Airport, Minnesota (FCM); and Centennial Airport, Colorado (APA). Crystal Airport had two scenarios, SOD and MIC. For each airport, researchers replicated one or two specific intersections identified as runway incursion hotspots and used them as the basis for a taxi scenario in the simulator. Objective and subjective data were collected throughout the simulation. Researchers analyzed the data to look for any patterns that are suggestive of runway incursion causal factors. Researchers viewed video tapes of pilots performing taxiing operations. The data collected included: total taxi time, taxi speed, number of stops, time spent looking at airport surface map, whether or not a surface incident occurred, scan time, and head-up time.

Apparatus

The original AMB study was conducted using a real-time, high fidelity general aviation cockpit simulator at the William J. Hughes Technical Center Cockpit Simulation Facility (CSF), configured as a Cessna 421. The visual system was a projector-based display system designed to provide the pilot/copilot with an Out of the Window (OTW) display on the windscreen. Three high resolution projectors were used to project the OTW view. Their purpose was to display a scene with realistic depth of field cues for the pilot. Microsoft Flight-simulator 2002 was used to generate and display the visual scenes. In addition, the audio system allowed for radio communication between pilot and controller, and provided simulated engine sounds.

Participants

28 pilots participated in the simulation study. The pilots were all General Aviation pilots, and had an average age of 43 years. They had an average of 13 years flight experience. Participants had logged an average of 1400 hours total flight time, and an average of 82 hours logged in the past 6 months.

Results

The results attempted to look at different aspects of pilot performance and behavior to see if there are any links to RIs/SIs. Statistical analyses were performed on the data and the results of these analyses are reported below. Communications data analysis is not reported here due to loss of sound in over 50% of the videotapes.

Out of 140 experimental runs in the simulation, there were 13 SIs, committed by 10 pilots. Three of these pilots committed two SIs. These were all violation of hold-short instructions.

The 10 pilots who committed SIs were on average less experienced than the other 18 pilots who did not commit SIs. In terms of overall flight experience, pilots who had SIs had logged fewer flight hours than pilots without incidents (482 flight hours versus 1940 hours). However, the test showed that this did not reach statistical significance $F(1, 25) = 3.956$, $p = .058$, as shown in Figure 1; therefore, the null hypothesis could not be rejected.

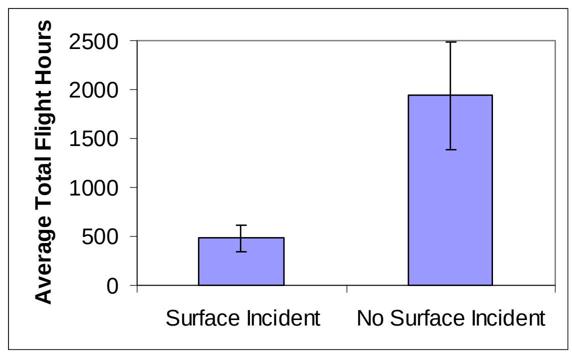


Figure 1. *Surface incidents as a function of overall experience level*

There was a statistically significant difference in number of flight hours in the past six months between those pilots who had at least one SI and those who did not (32.7 hours vs. 97.9 hours); $F(1,25) = 7.213$, $p < .05$, as shown in Figure 2. The recent flight hour data of one participant was excluded from analysis because the participant was also a commercial pilot who had many more flight hours than any of the other participants.

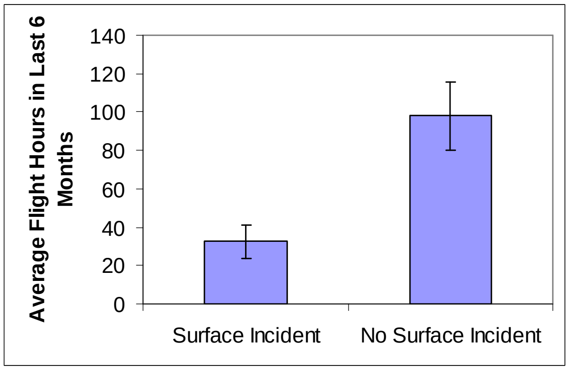


Figure 2. *Surface incidents as a function of recent flight time*

On average, those pilots with the lowest number of flight hours in the past six months had more SIs than those with higher numbers of recent flight hours; $F(2,24) = 3.578$, $p < .05$ (see Figure 3).

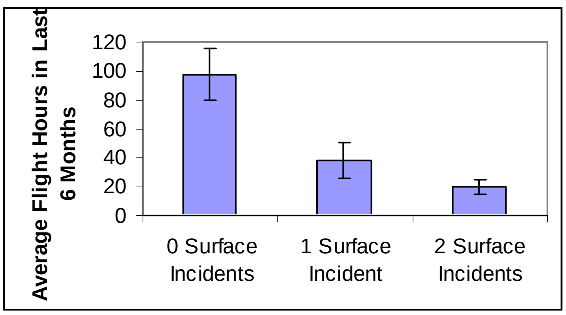


Figure 3. *Number of surface incidents committed by pilots as a function of average flight hours*

It was found that on average, pilots who committed SIs reported a smaller percentage of their flight experience at towered airports than those pilots that did not have SIs; $F(1,25) = 6.438$, $p < .05$. Those who committed SIs estimated that 26.6% of their flight operations were conducted at towered airports, while those who did not commit SIs estimated conducting of 55.1% of their flight operations at towered airports.

The amount of 'head-up' time spent in the scenarios show that those who committed SIs tended to spend less time looking out the window than those who did not commit surface incidents. However, there was not a statistically significant difference in overall percentage of head-up time between those who committed SIs and those who did not. The average percentage of time spent looking out the window by those who committed SI was 93.9%, while average head up time was 94.4% for those who did not commit an SI.

There was a significant difference in pilot-reported SA between scenarios, $F(4,24)=3.026$, $p < .05$. There was also a trend of lower SA with increasing numbers of SIs across scenarios (see Figure 4), with the exception of the LGB scenario. The LGB scenario had the second lowest SA rating, but only had one SI.

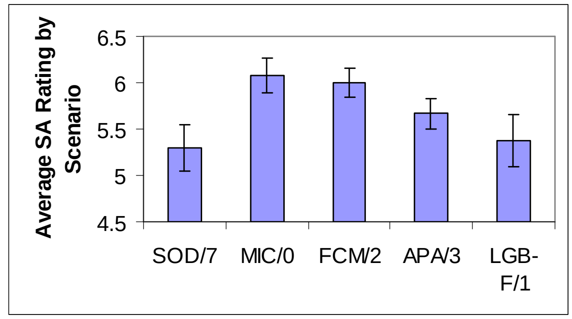


Figure 4. *Average situation awareness ratings as a function of number of surface incidents*

Pilots who committed SIs reported significantly lower SA than pilots without SIs; $F(9, 18) = 4.165$, $p < .05$ (see Figure 5).

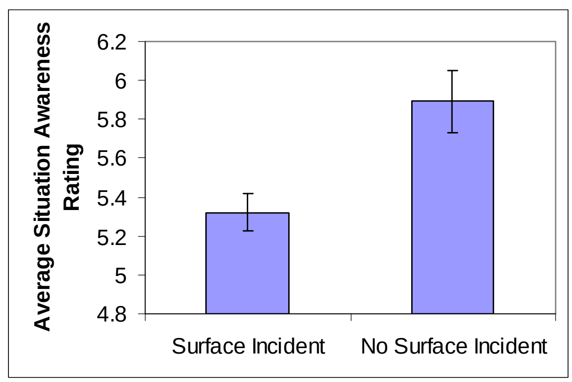


Figure 5. *Average situation awareness ratings as a function of surface incidents*

When looking at situation awareness ratings by scenario, there were some significant differences, as shown in Figure 6. In the MIC SOD scenario, those pilots who had an SI reported significantly lower SA than those pilots who did not have an SI; $F(1, 250) = 14.306$; $p < .01$. In the LGB scenario, pilots who had SIs also reported significantly lower SA; $F(1,25) = 14.389$; $p < .01$. In the FCM scenario, pilots who committed SIs tended to report lower SA. However, the difference was not statistically significant: $F(1,250) = 3.902$; $p = .059$.

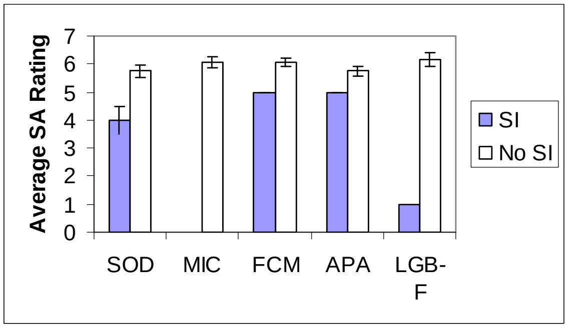


Figure 6. *Average situation awareness ratings as a function of scenario and occurrence of surface incidents*

Pilots who committed SIs tend to report higher mental workload (as measured by the NASA TLX) than pilots without incidents. However, the difference in reported workload ratings for those who had SIs and those who did not was not statistically significant: $F(1, 25)=3.205$, $p=.08$. The average reported workload for those participants who committed an SI was 47.78, and for those who did not commit an SI it was 40.196 (see Figure 7).

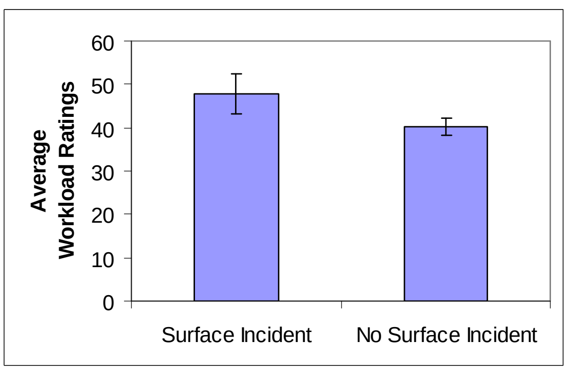


Figure 7. *Average workload ratings as a function of surface incidents*

No other data from the simulations indicated statistically significant differences. This data included the number of times pilots stopped during taxi, how long pilots spent looking at the taxi diagram (both before and during taxi), the rate of taxi speed, the number of times pilots scanned outside the cockpit and the amount of time spent scanning.

Discussion

This study examined pilot and scenario characteristics that may aid in pointing out causal factors for runway incursions and SIs. The results of this study suggest that there are some pilot characteristics that may be associated with a higher probability of committing SIs. Those pilots with fewer flight hours total tended to be more likely to commit SIs. Pilots with fewer flight hours in the past six months also tended to have more SIs. Pilots who spent less time at towered airports were more likely to have SIs. The amount of time spent looking out of the cockpit was found to be related to occurrence of SIs, but was not enough to be used as a predictor of SIs. Those pilots who commited SIs tended to report lower levels of SA, and higher levels of workload than those pilots who did not commit SIs.

The finding of a significant difference in flight hours during the past 6 months suggests that pilot training/experience may have an effect on the rate of SIs. The results suggest that the more recent flight experience a pilot has, the less likely the pilot is to commit an SI. Logging more recent flight hours aids the pilot in maintaining proficiency with proper procedures and pilot skills. The amount of total flight time and experience a pilot has is also related to whether or not they had a SI. Those pilots who had a SI tended to have fewer total flight hours than those who did not have a SI. Taken with the findings on recent flight hours, this suggests that experience and training may be predictive which pilots are more likely to commit an SI.

Pilots who committed SIs reported a lower percentage of their flight operations being conducted at towered airports. This finding suggests that pilots who do not have much experience at towered airports are more likely to commit an SI. The larger size and higher complexity of towered airport layouts may contribute to pilots who are not used to the size and complexity being more likely to commit an SI. In addition, the increased amount of traffic and the need to communicate with ATC may also increase the workload of pilots not used to operations in the towered environment. The increased workload may reduce the SA of pilots who are not as familiar with towered airports.

There appears to be a trend relating reported SA and scenarios. Pilots generally reported lower SA in the scenarios with higher incident rates. Although all of the scenarios were chosen for the simulation because they were known to have a high frequency of pilots committing SIs, the pilot participants had more difficultly with certain scenarios than others, as evidenced by both the higher frequency of SIs committed, and the lower average SA ratings for those scenarios.

Overall, pilots who committed an SI had lower SA than pilots who did not. In addition, although the difference was not statistically significant, pilots who committed an SI had higher average workload ratings. It is likely that the combination of less recent flight experience and less experience at controlled airports were the cause of increased workload and lower SA for some pilots. The fact that the pilots were taxiing for the first time at airports known for high rates of SIs also likely led to increased workload and lowered SA. For example, the MIC SOD scenario instructed pilots to hold short of a sod runway. This is an unusual element on the airport surface, and pilots may not have had the necessary experience with it to know what to look for. As a result, there was lowered SA, seemingly attributed to the airport surface. The resultant increased workload and decreased SA led to a higher likelihood of these pilots committing SIs.

The fact that those pilot who had higher levels of recent flight hours also tended to not have SIs suggests that experience and/or training may help in reducing incidents. The recency of experience may also be predictive of surface safety. Continued training, and keeping pilots current with flight hours may be a way to help reduce SIs.

While there were some factors identified from this research that are suggestive of predictors of who will commit an SI, the study was not designed with this in mind. In order to confirm the predictive factors of who will commit SIs, it would be necessary to design a study with that purpose in mind.

Future efforts to continue to examine runway safety should include the development of a model to synthesize these results. The model may be used to predict SIs, thereby generalizing the results to other airports and scenarios, as well as confirming and identifying additional pilot characteristics that increase the likelihood of SIs. Therefore, current knowledge of the airport surface and pilot characteristics could be used to more accurately predict and reduce the likelihood of SIs.

References

Adam, G.L., Lentz, R.H., & Blair, R.W. (1992). Study of the causes of runway incursions and related incidents. *IEEE*, 1992.

Air Traffic Evaluations and Investigations Staff (2002, 15 August). *Air Traffic Quality Assurance*. Washington, D.C.: Federal Aviation Administration.

Jones, D.R. (2002). Runway incursion prevention system simulation evaluation. *Presented at the 21st Digital Avionics Systems Conference*, Irvine, CA, October 27-31, 2002.

Koenig, R.L. (1995). FAA report surveys U.S. airline pilots to discover factors that promote runway incursions. *Airport Operations*, July-August 1995.

Off ice of Runway Safety, ARI (2001, June). *Runway incursion trends at towered airports in the United States FY 1997-2000*. Washington, D.C.: Federal Aviation Administration.

Office of Runway Safety, ARI (2003, June). *Runway incursion trends at towered airports in the United States FY 1999-2002*. Washington, D.C.: Federal Aviation Administration.

Office of Runway Safety, ARI (2004, August). *Runway incursion trends and initiatives at towered airports in the United States, FY 2000 – FY 2003*. Washington, D.C.: Federal Aviation Administration.

Pope, J.A. (1990). The airport side of runway incursions. *Flight Safety Foundation, Airport Operations*, Sept/Oct 1990.

Surface Technology Assessment Product Team, AND-520 (2004, January). *Addressable Message Board simulation study final report*. Atlantic City, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Technology Assessment Team (2002, December). *Runway incursion airport assessment report*. Washington, D.C: Federal Aviation Administration.

Young, S.D., and Jones, D.R. (1998). Flight testing of an airport surface guidance, navigation, and control system. *ION National Technical Meeting*, January 21-23, 1998.

SUPPORTING TEAMS IN CRISIS WITH IT: A PRELIMINARY COLLABORATION FRAMEWORK

Earl McKinney Jr.
Bowling Green State University
Bowling Green, Ohio

To support their teams in crisis, organizations seek to leverage advances in information technology. These advances include *automation* support to the warfighting team (e.g. an electronic checklist for a flight crew), as well as *collaboration* support such as linking engaged combat troops to intelligence services. While automated support is rapidly developing, very little consideration has been given to enhancing the collaboration support for teams that face crisis. Here we suggest a preliminary set of IT system attributes to support collaboration for teams that face crisis. These attributes are based on two frameworks that have been developed to mitigate the effects of crisis. One is an organizational approach called the High Reliability Organization (HRO), the other, a team approach based on Crew Resource Management (CRM). Here we suggest attributes of an IT system to support teams that face crisis based on these two approaches.

Introduction

To support their teams in crisis, organizations seek to leverage advances in information technology. These advances include *automation* support to the warfighting team (e.g. an electronic checklist for a flight crew), as well as *collaboration* support such as linking engaged combat troops to intelligence services.

Understanding how to support teams that face crisis is essential. Currently IT support includes display systems (Hamblin, 2003; Sarter & Schroeder, 2001; Vicente, 2003), intelligent support systems (Koester & Mehl, 2003; Palmer & Degani, 2001; Wischusen et al., 2003), decision support systems (Smith, Johnson, & Paris, 2004), and a wide variety of other technical solutions (Stoner, et al., 2004). These systems give automated support to teams in crisis. However, very little consideration has been given to enhancing the collaboration support for teams that face crisis (Huang, 2004; Nunamaker, 1997).

Collaboration

Collaborative support for these teams in the past was limited by the available technology. Historically, flight crews, military teams, or surgical units could not be collaboratively supported as only the team in crisis knew the local conditions and had access to the stand alone computers that produced the crisis data. In the past, teams in crisis had only their immediate resources at hand or preprogrammed automated support. Now, with advances in network capacity and sensors, IT has stretched that hand and teams that face crisis can obtain collaboration support from others in the organization. These organizational experts can now see real-time data from the crisis, interact with knowledge bases, and reliably communicate with the team.

To date, IT support for crisis teams has focused exclusively on automated support. Teams are supported with a variety of tools such as electronic checklists, self contained expert systems, and agent technology. However, we suggest one fundamental principle of crisis is that it is unexpected, unpracticed, and unprogrammable (McKinney & Davis, 2003). Whereas an engine failure or low oil temperature on an engine may be an emergency, examples of crisis include being shot, responding to a novel terrorist attack, facing an engine failure over a combat zone, or responding to novel combinations of systems failures. Emergencies are predefined and therefore amenable to automated support. With an emergency, we know what is wrong and we can optimize and train a specific response and support that response with automated support such as a checklist or sensor or display device. Crisis by its uniqueness reduces the utility of automated support. The challenges are figuring out what is happening, and thinking through irrevocable decisions. As a result, automated support while valuable should not be the only available support for teams that face crisis. Collaboration with other human experts is necessary to aid problem discovery and to consider ramifications of responses.

Teams that Face Crisis

As an organizational component, crisis teams inherit the organization's resources, culture, and goals. Characteristics of the organization have been shown to have a significant effect on crisis team performance. For example, organizational culture has been shown to affect team performance (Bierly & Spender, 1995), and organizational goals and strategy also significantly impact team behaviors (Kozlowski, 1998). If team performance is strongly affected by organizational aspects, a framework to support teams in crisis should be based in part on organizational activities that support these teams.

most extensive insight into organizational activities that mitigate the effects of crisis has been labeled High Reliability Organizations (HRO). Weick and Sutcliffe (2001) originated the HRO framework. They trace the success of organizations that have teams that face crisis to five activities. These include preoccupation with failure, reluctance to simplify interpretations, sensitivity to operations, commitment to resilience, and deference to expertise. The first three reduce crisis incidence while the last two enhance resilience. More detail on these five activities will follow. In later sections, attributes of the collaborative support system will be organized under these activities. Supporting crisis teams with IT should be based on these five "team-organizational" activities.

While support for these team-organizational activities is important to crisis team success, it is also valuable to consider what might, by contrast, be labeled team only needs. The activities of teams in crisis have been the object of flightdeck research for 25 years. This research effort, labeled Crew Resource Management (CRM) suggests that team-only needs might include situational awareness, decision making, communication, team work, personal resources and leadership.

Here, we combine these two models and present IT principles to support collaboration needs using both the HRO team-organization activities and the CRM team activities. Due to space limitations, we only explain the HRO activities in depth as CRM activities are more familiar to this audience.

IT System Attributes

The following list of IT system attributes is based on a review of the HRO and CRM activities. An explanation of the HRO activities and their corresponding system attributes are further discussed after the list.

System Attributes Based on HRO Activities

1: Encourage widespread near miss and error reporting and analysis that lead to improved processes
2: Permit recording of detailed accounts of near misses or errors that allows new attributes to be collected and analyzed
3: Provide the opportunity to retain and display unsimplified data and disconfirming evidence
4: Track and display a wide variety of data for a variety of expert interpretation
5: Increase the visibility of operational performance measures and reward operational enhancements that lead to continual improvements
6: Reward operational change and adapt to changes in operations
7: Allow simultaneous action and diagnosis while supporting on going activities
8: Permit depth of analysis and mental simulation of courses of action
9: Identify and match experts with on going problems
10: Supports ad hoc team communication and analysis among experts

System Attributes Based on CRM Activities

11: Be simple--don't overly filter or over process the original data, just put the data into meaningful form
12: Help reduce mental effort by supporting feature matching and story telling
13: Display information cues and historical trends in such a way that the load on an operator's short term memory is minimized
14: Provide a mechanism to direct the attention of an operator to important events minimizing the cognitive costs of interruption
15: Provide a mechanism to mitigate the effects of confirmation bias
16: Suggest actions that would provide diagnostic feedback from situations in which information cues are equivocal, thereby mitigating the tendency to attend only to the information we want to believe
17: Compensate for deficiencies in action selection (what to do about it)
18: Enable communication value sharing
*19:*Aid increased vertical communication during crisis
*20:*Support communication of effective dissent or alternative hypotheses
*21:*Enhance accuracy and sharing of common models on the state of affairs

Principles of HRO-Organizational Activities and System Attributes

In the following sections, the activities of successful HROs are outlined. Within each section, the attributes of an IT system to support each activity are also presented. Examples of successful HROs from the aviation industry are also included. Descriptions of HRO activities are based on *Managing the Unexpected* by Karl Weick and Kathleen Sutcliffe (2001).

1. Preoccupation with failure

Members of HROs constantly worry about failure and distrust success. They look hard for lapses or minor incidents that, if ignored, could later recur and lead to significant failures. This preoccupation with failure is impervious to success. HROs distrust success as it tends to narrow perception and breed overconfidence. This misplaced confidence in judgment and in existing procedures limits changes to the organization and its processes. One way HROs fight the lethargy of success is by establishing attribution-free error reporting procedures. Anyone in the organization can report errors and are assured that those errors will not lead to sanction. These error reports are never automatically or thoughtlessly processed by the HRO. Rather the data collected is turned into active incident reviews and in depth analysis that are widely communicated.

A manifestation of preoccupation with failure in the airline industry is error reporting (Chidester, 2003). The Aviation Safety Reporting System (ASRS) is one national system, and all major airlines have their own internal systems. Pilots make inputs to the systems via anonymous reports (see ASRS at http://asrs.arc.nasa.gov/). Data from these systems are then analyzed by trainers and researchers. Their reports are widely shared and the results of the studies have had significant impacts (Gunther, 2003). The ASRS is just one of several examples of airline preoccupation with failure. Training departments at airlines continually develop new error frameworks and mitigation processes (Chidester, 2003). Further, they are increasingly active in analysis and communication of errors and abnormal situations (Haney & Gertman, 2003; Muthard & Wickens, 2003).

System attribute 1:

Encourage widespread near miss and error reporting and analysis that lead to improved processes

These near misses and errors may contain warnings of future problems but in the din of daily activity appear as only weak signals of impending crisis. The IT system must be designed to find and amplify these weak signals. Unfortunately, weak signals, by their nature, are not readily found as they defy easy classification or categorization. If categories or attributes of errors were already known to the organization, the errors that occur would also be known and procedures established to respond. For example, jet engines break down, and therefore airlines have learned to classify these failures as engine problems. However, most weak signals are not easily classified (e.g. how should a small crack, or mistyped clearance be classified?). As a result, most organizations can not respond until the wing crack leads to a break and a crisis occurs. Thus, the crisis IT system should permit detailed descriptions or detailed reporting of odd events, near misses, and weak signals. From these details, common attributes, such as the length of a "must repair" crack, or the frequency of clearance errors can later emerge. Once these new attributes are known, tolerances can be set for future inspections and reporting and attention can shift to finding new attributes or categories.

System attribute 2:

Permit recording of detailed accounts of near misses or errors that allows new attributes to be collected and analyzed

2. Reluctance to simplify

High reliability organizations refuse to simplify the complex events in which their teams participate. Although all coordination requires some simplification, in HROs, participants minimize this simplification. Instead, they constantly seek to see more, and render more complete and detailed their understanding of both their actions and the environment. When actions are taken they avoid the common simplifying process of seeking confirming evidence that their actions were appropriate. Rather, they seek disconfirming evidence that expectations and experience can conspire to hide.

System attribute 3:

Provide the opportunity to retain and display unsimplified data and disconfirming evidence

One way HROs generate disconfirming evidence for their teams is by assigning members with varied and overlapping backgrounds to the crisis team. The variety in backgrounds tends to increase the data that are scrutinized and thereby increase the variety of what can be noticed. By creating teams with members who have overlapping experiences the team is able to see a more complete perspective on their actions and the environment. In addition to variety in the team members, one other source of variety is organizational expert variety. The crisis IT system brings this variety of organizational experts online with the crisis team, allowing them to notice, to suggest, and to think ahead with those in the crisis. With varied backgrounds comes varied experiences and expectations and skepticism of simplification. In addition to the variety of the team, and variety of organizational experts, the search for disconfirming evidence is also enhanced by a varied search of a wide variety of sources. Therefore, an IT system that limits simplification would have a variety of sensors that records a variety of data for a variety of participants.

System attribute 4:
Track and display a wide variety of data for a variety of expert interpretation

3. Widespread sensitivity to operations
HROs value operations above strategy. This focus on current operations is designed to find hidden or underlying lessons about weaknesses in the operation. These latent failures may be found in many areas including poor supervision, inadequate procedures, and deficient training. In addition to finding and correcting these significant operational failure points, HROs also demonstrate their commitment to operations by their focus on correcting even minor issues. The result is continuous improvement in operations. To sustain this incremental improvement, HROs seek operational suggestions from the whole organization. They widely disseminate and seek feedback on both operational performance and performance measures. This operational priority is evident in other ways-- in the attention devoted to even small interruptions in operations, in the frequent meetings on operational status, and in organization structure designed to widely distribute real time information about operations.

Airlines are an example of HROs committed to operations. At major hub airports, airline ground support centers demonstrate this sensitivity to operations. These centers refuel, clean, restock, and support all passenger and flight activities at the hub. Operational performance of the hub is closely tracked and widely disseminated throughout the company. For example, the on-time departure percentage of the first set of flights leaving the hub are calculated and compared to benchmarks and to other hubs at the airline. Every hub center knows how they compare real time to other hub operations. They work collectively to constantly refine gate allocation algorithms, refueling procedures, and clearance conflicts to continually improve operational measures such as on-time performance and resource use.

Not only should operational performance data be available for local use, IT systems supporting teams in crisis should be designed to widely disseminate the state of current operations within the organization. The system should make operational data, training schedules, and other process information increasingly available for oversight and improvement. This should result in improvements to operational procedures from a variety of sources.

System attribute 5:
Increase the visibility of operational performance measures and reward operational enhancements that lead to continual improvements

One key implication of operational process change is that the IT system itself must change. Therefore, the system must be flexible enough to adapt to changes to operations.

System attribute 6:
Reward operational change and adapt to changes in operations

4. Commitment to resilience

HROs are built on the premise that error is inevitable. As a result, HRO managers take pride in engaging in putting out fires. Unlike managers in other organizations who see fire fighting as a failure of planning and a drain on resources, HRO managers know that recovery from error is their primary activity. Because of this priority they seek deep knowledge of their technologies, processes and people. In addition, they excel at adapting to swift feedback, learning quickly without error, recombining existing responses, and mentally simulating courses of action. Further, they have learned to treat while diagnosing and to adapt to threats based on feedback from action.

The professional aviation community has realized that error is inevitable. In fact, one report estimates the frequency of pilot error at 5-10 mistakes per hour (Amalberti, 1996). As a result, flight systems, training, technical systems, and procedures are designed to respond and recover from emergencies. Further, pilots are taught detailed knowledge about their aircraft systems, and their environment in order to more accurately diagnose crisis and think through courses of action.

System attribute 7:
Allow simultaneous action and diagnosis while supporting on going activities

System attribute 8:
Permit depth of analysis and mental simulation of courses of action

5. Deference to expertise
As implied earlier, HROs deliberately employ a wide variety of expertise to avoid simplification when responding to crisis. Not mentioned earlier is how

HROs are organized to deploy that expertise. Expertise is not employed in a rigid organizational structure, rather experts are expected to self organize around a problem. In addition, they are permitted to make changes without multiple levels of oversight common in more hierarchical organizations. By pushing responsibility and authority down and out to where the organization meets its environment errors are noted earlier and problems more quickly addressed. Moreover, operating dynamics are such that when the signals emanating from the crisis are noticed, experts find the problem and resolve it at a low level. Quick and accurate decisions by those closest to the action are emphasized. Westrum call this coordinate leadership (Westrum, 1997).

Currently airlines provide a poor example of deference to expertise. Aircraft operational decisions are vested in the captain, and the crew, with only limited support from other organizational experts. While crews can use their two way radio to ask for maintenance or weather support, the crew is cut off from other experts in the organization and is alone responsible for finding all potential problems. A better system would allow crews to have on going collaborative support that during a crisis would grow to include a number of company experts.

To support better use of expertise the IT system for teams in crisis must permit data and analysis to migrate to appropriate experts. It should encourage signal watchers close to the action to alert the right experts in the organization about anomalies. As a result, exception reporting, and other signals of problems should not just go to executives but be shared widely within the organization.

System attribute 9:
Identify and match experts with on going problems

In addition, the IT system must be configurable to these ad hoc collaboration teams. In contrast to supporting these ad hoc teams, traditional IT systems have the effect of making organization decision making rigid and predefined. However, the goal for a crisis system should be to support the analysis needs of a variety of experts in ad hoc teams.

System attribute 10:
Supports ad hoc team communication and analysis among experts

Summary

To date, little work has investigated supporting the collaborative needs of teams that face crisis. The uniqueness of the crises event suggests that in addition to automated support, teams that face crisis would benefit from real time collaboration from other experts in the organization.

The goal of this report was to develop an initial list of IT system attributes to support teams in crisis. To accomplish this, two main frameworks of crisis were reviewed. The first model, High Reliability Organizations, suggests that to mitigate the effects of crisis teams should be preoccupied with failure, avoid simplifications, attend to operations, commit to resilience, and defer to expertise. The second, Crew Resource Management (CRM) posits that effective decision making, communication, and a shared situational assessment contribute to an effective response to crisis. Using these eight activities, 21 specific and distinct attributes of a crisis IT system were presented. Future research should further refine this list, evaluate its completeness, and assess its generalizability. As with other studies of crisis, it is difficult to collect observations or conduct experiments. On the other hand, as cockpit voice recorders and flight data recorders become more common, more scientific analysis of the system attributes suggested here will be possible.

References

(Full list available from author)

Bierly, P. E., & Spender, J. (1995). Culture and High Reliability Organizations: The case of the nuclear submarine. Journal of Management, 21, 4, 639-656.

Cohen, S. (1980). Aftereffects of Stress on Human Performance and Social Behavior: A Review of Research and theory. Psychological Bulleting, 88, 82-108.

Davis, R., Driskell, J.E., & Salas, R. (1991). Group decision making under stress. Journal of Applied Psychology, 76, 473-478.

Driskill, J., Salas, E., & Johnston, J. (1999). Does Stress Lead to a Loss of Team Perspective? Group Dynamics, 3, 1-12.

Esterbrook, J. A. (1959). The Effect of Emotion on Cue Utilization and the Organization of Behavior. Psychological Review, 66, 183-201.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. Journal of Abnormal and Social Psychology, *5*, 203-210.

Hirokawa, R., & Poole M. (Eds.) (1996). Communication and group decision making. Thousand Oaks, CA: Sage.

Huang, Q. (2004). "Embedded Simulation for Decision Support." Proceedings of 13th Conference on Behavior Representation in Modeling and Simulation, Arlington VA, 399-400.

Jehn, K. (1999). A multimethod examination of the benefits and detriments of intragroup conflict. Administrative Science Quarterly, 40, 256-282.

Matthews, K., & Cannon, L. (1975). Environmental Noise Level as a Determinant of Helping Behavior. Journal of Personality and Social Psychology 32, 571-577.

McKinney, E., & Davis, K. (2003). "The Effects of Deliberate Practice on Crisis Decision Performance." Human Factors, Winter, 45, 398-416.

McKinney E., Barker, J., Smith, D., & Davis, K. (2004). "The Role of Communication Values in Swift Starting Action Teams: IT Insights From Flightcrew Experience." Information & Management, 41, 8, 1043-1056.

McKinney, E., Barker, J., Smith, D., & Davis, K. (2004). "How Swift Starting Action Teams Get off the Ground: What Airline Flightcrews can tell us about Team Communication." Unpublished manuscript, Bowling Green State University.

Morrison, J., Kelly, R., Moore, R & Hutchins, S. (2000). "Implications of decision-making research for decision support and displays," in J. A. Cannon-Bowers and E. Salas (Eds.), Making Decisions under stress: implications for individual and team training. Washington DC: APA books.

Nunamaker, J. (1997). Future research in group support systems, International Journal of Human Computer Studies, 47, 3, pp. 357-385.

Olson, W., & Sarter, N. (2001). Management by consent in human-machine systems: When and why it breaks down. Human Factors, Summer 2001, 43, 255-266.

Orasanu, J. M. (1990). *Shared mental models and crew decision making.* (CSL Report No 46). Princeton NJ: Princeton Univ.

Rosenthal, U., & 't Hart, P. (1991). Experts and decision makers in crisis situations. Knowledge: Creation, Diffusion, Utilization, Vol 12, 4, June, 350-372.

Rotton, J., Olszewski, D., Charleton, M., & Soler, E. (1978). Loud speech, Conglomerate Noise and Behavioral Aftereffects. Journal of Applied Psychology, 63, 360-365.

Seeger, M., Sellnow, T., & Ulmer, R. (2003). Communication and Organizational Crisis. Westport Connecticut: Praeger

Smith, C. A. P., Johnston, J., & Paris, C. (2004). Decision support for air warfare: Detection of deceptive threats. Group Decision and Negotiation, 13, 129-148.

Te'eni, D. (2001). Review: A cognitive affective model of organizational communication for designing IT. MIS Quarterly, June, pp., 251-313.

Tolcott, M. A. (1992). Understanding and Aiding Military Decisions. Inn S. C. Collyer (Ed.) The 27th International Applied Military Psychology Symposium: a focus on Decision Making research (ONR Europe Report 92-4 W, pp. 33-48). London: Office of Naval Research.

Vicente, K. (2003). Less is (sometimes) more in cognitive engineering: the role of automation technology in improving patient safety. Quality and Safety in Health Care, August, 12, 291-295.

Walton, R. E., & McKersie, R. B. (1965). A Behavioral Theory of Labor Negotiations: An Analysis of a Social Interaction System. New York: McGraw Hill.

Weick, K. & Sutcliffe, K. (2001). Managing the Unexpected, San Francisco CA: Jossey Bass.

Weick, K. E. (1990). The Vulnerable System: An Analysis of the Tenerife Air Disaster. Journal of Management, 16, 571-593.

Westrum, R. (1997). "Social Factors in Safety Critical Systems." In Human Factors in Safety Critical Systems, Redmil, R. and Rajan J. (Eds.) London: Butterworth-Heinemann.

A FRAMEWORK FOR THE ASSESSMENT OF CONTROLLER COORDINATION IN THE AIR TRAFFIC CONTROL TOWER ENVIRONMENT

Peter Moertl
The MITRE Corporation
Center for Advanced Aviation System Development (CAASD)
7515 Colshire Drive M/S N420
McLean, Virginia 22102-7508
Voice: (703) 883-1080
Fax: (703) 883-3348
Email: pmoertl@mitre.org

Craig Bonaceto
The MITRE Corporation
202 Burlington Road Rte. 62 M/S M370
Bedford, MA 01730-1420
Voice: (781) 271-3866
Fax: (781) 271-2101
Email: cbonaceto@mitre.org

The air traffic control tower (ATCT) environment requires coordination between various controller positions for safe and efficient operations. With the development of new collaborative decision support tools in the tower (e.g. SMS, NASA 2004), efficient human interface design will require the consideration of the coordination routines that controller use. Whereas inter-position coordination is generally prescribed by the FAA (FAAO 7110.65) as well as by specific ATCT standard operating procedures, little is known about face-to-face coordination that is not captured by other recording media such as flight strip marking or radio communication.

To meet the demand for more information about air traffic control tower coordination, a framework for ATCT coordination was developed in cooperation with operational experts. Using a card sorting technique, ATCT controllers ranked various ATCT coordination events that had been identified by Alley et al. (1987), commented on their experience with each coordination event, and added additional coordination events. From these comments, relevant coordination dimensions were extracted that included coordination frequency, mental workload, coordination time criticality, as well as environmental factors that influenced the coordination. Controllers then quantified their experience using these dimensions to confirm and modify the framework. The proposed coordination framework is intended for the assessment and quantification of coordination in specific ATCT environments.

BUILDING AN INTERLOCKING HUMAN FACTORS RESEARCH AND DEVELOPMENT PROGRAM

Jacqueline Rehmann
Federal Aviation Administration Technical Center
Atlantic City, New Jersey

Dino Piccione
Federal Aviation Administration
Washington, DC

Richard Mogford, Ph.D
National Astronautics and Space Administration
Moffet Field, California

The Federal Aviation Administration (FAA) partners with the National Astronautics and Space Administration (NASA) to manage and integrate research on enhanced air- and ground-based air traffic management technologies. This partnership, designed to integrate air traffic decision support tools, concepts, and procedures, was formalized in September 1995. Coordinated research initiatives are described in joint research project descriptions (JRPDs) that define objectives, approach, responsibilities, mission relevance, goals, and outcomes. JRPD 12 is unique in that, as a cross-cutting JRPD, it ensures relevant human factors research issues, methods, metrics, and findings of individual programs of both organizations are made known to, shared, and leveraged by the larger research community including FAA, NASA, aviation industry and academia. Now, FAA is challenged as never before to integrate research and development (R&D) capabilities into the National Airspace System. The lessons and challenges identified by this group are summarized and presented as recommendations for establishing an integrated and focused human factors R&D program.

Introduction

The Federal Aviation Administration (FAA) and the National Astronautics and Space Administration (NASA) have partnered through the Interagency Air Traffic Management Integrated Product Team (IAIPT). This partnership, designed to integrate research addressing air traffic management (ATM) decision support tools, concepts, and procedures, was formalized in September 1995. Oversight for IAIPT activities is provided by the FAA's Research, Engineering, and Development (R,E,&D) Advisory Committee and NASA's Aeronautics and Space Transportation Technology Advisory Committee. The IAIPT is intended to ensure that shared research provides new technologies, procedures, and concepts of use for the National Airspace System (NAS).

The FAA is responsible for the installation, operation, and maintenance of the NAS. Research organizations such as NASA, MITRE CAASD, and the Massachusetts Institute of Technology/Lincoln Laboratories generate technologies and concepts for the future NAS that are candidates for implementation. The implementation process often makes use of commercial vendors that provide the transition from research prototypes to production units that can be installed and supported at any facilities operated by the FAA. It is essential that there is an open and harmonious line of communication between the various organizations. The IAIPT provides a forum for information exchange and assures that there is a means to share the vision for the future NAS to meet capacity and safety goals and smooth the path from research to acquisition.

The IAIPT manages the pipeline for how maturing research concepts and prototypes flow into the FAA acquisition management system. The emphasis is on the maturation of capabilities from laboratory to field implementation using a model of Technology Readiness Levels (TRL) to coordinate objectives, outputs, and exit criteria for moving from one level to another. As research proceeds, the TRL model provides milestones to ensure organizations increasingly specify the capability. The TRL paradigm describes an ideal where research is matched by an orderly transfer process into implementation. In fact, this migration is not always adhered to for a variety of reasons. Now more than ever, as budget cuts and prior obligations complicate FAA's ability to incorporate research into NAS modernization, our efforts and those of our research partners need to be leveraged. As our research partners continue to dedicate resources for air traffic management research and development (R&D), the challenge is to ensure that all ATM R&D moves toward a common vision for the future NAS.

Discussion

IAIPT research initiatives are described in joint research project descriptions (JRPDs). JRPD 12 ensures that relevant human factors research issues, methods, metrics, and findings of individual programs of both organizations are shared with the larger research community. Specifically JRPD 12 is intended to provide "a framework to systematically identify, coordinate, and integrate human factors efforts in the research and development of advanced ATC/ATM/CNS automation, technologies, concepts and procedures."

For the past 6 years, human factors practitioners from FAA, NASA, and various organizations have met to exchange research findings and lessons learned. This group generally focuses on specific research topics, and findings, issues, challenges, and lessons learned are shared. The exchange is intended to help researchers avoid the problems of the past and identify areas where future research could bear fruit. Technical information meetings are an important way for IAIPT members to exchange information and perspectives. Participants discuss the important contribution of human factors in transitioning research concepts and products through the R&D pipeline to acquisition and fielded systems. Meeting participants have generally agreed that transitioning research concepts from exploration, development, and acquisition of fielded systems should be accompanied by increasingly detailed assessments of information requirements, display management and integration, human centered automation, and human performance assessments that measure workload, situation awareness, and human error. All agree that human factors assessments are part of a larger integrated system engineering perspective encompassing operational concepts, system requirements, and system engineering methods.

More recently, discussions have focused on the need to collaborate earlier in the research cycle to address the "business case" for changes in the NAS, including an interagency review of research to provide input on research intersections and value; and data sharing for model development and verification. Human factors practitioners have a role in the Air Traffic Organization (ATO) as a member of the FAA's Development Liaison Team (DLT) to assess the human factors aspects of proposed new capabilities. Management depends on human factors input for the business case to establish the return on investment for each candidate research capability, assess the likelihood that human performance will match system demands, and determine that safety goals will be met.

The DLT performs their assessment by reviewing the proposed technology as a potential contributor to the agency's goals for capacity and safety. The FAA is acutely concerned about the safety of the NAS; especially incidents involving the loss of separation between aircraft caused by human error (i.e., operational error), and will scrutinize each proposed technology to determine if it will affect the level of safety risk in the NAS. Many technologies attempt to provide decision support capabilities for air traffic controllers in an effort to increase capacity by recommending solutions to traffic problems that controllers must manage. Other technologies have the potential to drastically alter the role of the controller by using automation in various modes. In each case, the human factors representative on the team works with System Engineering, Technology Development, and FAA Technical Center representatives to provide Air Traffic Organization (ATO) management with recommendations.

The human factors issues are best assessed when there is human performance data to work with. The data helps human factors representatives understand how the technology will be integrated into the workstation, how it will affect procedures, and how the roles and responsibilities of humans in the system (operators and maintainers) will change. Human performance data should support the purported improvements in the NAS for both safety and capacity. The role of human factors in this context is to help FAA make informed decisions regarding investment. Our responsibility is to assess risk and provide insights regarding the impact of technology on human error. In addition, we must determine if the proposal will provide the level of service expected by the flying public as part of a national transportation system.

The DLT will work in concert with the IAIPT to provide guidance to the air traffic research community regarding the type of data that is needed. As technology matures through the TRL process, the developing organizations should be responsible for conducting the appropriate human factors activities during each stage of technology development (Krois, Mogford, & Rehmann, 2003). Once the decision is made to consider incorporating a technology into the NAS, the FAA's Acquisition Management System (AMS) provides the structure for eventual system procurement.

The JRPD-12 meetings are an opportunity for human factors researchers and practitioners in all member organizations to share information and concerns. The information sharing takes a number of forms including traditional technical presentations, programmatic presentations, guidance for navigating the waters of research and acquisition during periods of change, and clarification of goals and objectives that may lose meaning when organizational lines are crossed.

The latest J-12 Technical Information Meeting (TIM) covered a range of topics including such provocative topics as “Air Traffic Controller Staffing and the Age 56 Rule.” Other topics included Safety Management, En Route Research, NASA’s Air Traffic Management (ATM) research program to date, human performance and cognitive modeling.

A portion of the TIM was spent on the human factors aspects of the transition from a research program, through acquisition, into the fielding and daily use of a system. Researchers were provided with an opportunity to understand the perspective of human factors practitioners that need human performance data to address productivity, cost/benefit, staffing, skills, training, and human error. These human performance data are needed as essential input to trade-off and investment analyses that are used to determine if a system should be allowed to proceed to the next research level, or to enter the acquisition process.

Personnel and training costs are the largest contributors to the FAA’s operations budget and as new technology is introduced, there is a need to understand how productivity and staffing will be affected. In addition, the FAA is faced with a large turnover in the air traffic controller population In the next ten years. As new controllers are screened, trained, and assigned to new duties, the FAA needs a clear understanding of the number and types of individuals that will be needed to staff the air traffic system of the future. The human-system integration aspects of system design for the future NAS is a subject that requires research and analysis by the human factors community. While this was briefly discussed during the TIM, this will possibly be a topic for further discussion in a future meeting.

In addition to the topics covered, a panel of experienced human factors practitioners from several organizations was convened to consider a number of questions to help guide workshop discussions on the topic of “building an interlocking human factors ATM program.” The questions were:
-*What are the obstacles to building an interlocking human factors R&D program?*
-*What are the benefits?*
-*How would we proceed?*
-*What would be the characteristics?*
-*How would we achieve true collaboration?*
-*What have we learned from the past?*

Obstacles

During the discussion, attendees agreed that “collaborative” or “interactive” would be a better term to describe our organizational relationships. The initial discussion focused on obstacles to collaboration, such as organizational “stove-piping” which has hindered our ability to exchange information. Tight budgets also create competition for scare research dollars between researchers. And often, the momentum for collaboration is lost within 2-3 months of our meetings. This is exacerbated by a lack of ongoing communication between agencies, and a lack of visibility about research taking place in other organizations. Often, there is imperfect communication between researchers and sponsors. A lack of success criteria and continuously shifting priorities of research organizations further complicate the human factors R&D landscape.

Benefits

The benefits of collaboration are many, including preventing duplication of efforts and minimizing the cost of research. Technology is bringing potentially huge changes in roles and responsibilities for humans in the NAS. This is increasing, not decreasing, the need for collaboration. Performance of the NAS hinges on effective human performance. Moreover, the future NAS is predicted to need 2 or 3 times the capacity of the current system. Thus, human factors researchers will be even more challenged to approach problems from a human-system perspective and avoid piecemeal solutions. As safety and security concerns rise, we must assess the human component from a risk standpoint for any proposed changes to the NAS.

How to Proceed

How could a collaborative human factors research program work? Participants agreed that the FAA collocation study was a good start to examine the impact of multiple tools on the controller workstation. The study assessed the collocation effects of controller decision support tools that were developed independently by FAA, NASA, and MITRE CAASD. The tools included Traffic Management Advisor (TMA-NASA), Controller-Pilot Data Link (CPDLC-FAA), and User Request Evaluation Tool (URET-MITRE CAASD). The study identified important human factors issues that were not evident until the tools were expected to work together at a single controller workstation.

Several definitive steps were taken after the TIM to ensure our discussions about collaboration became reality. Selected FAA human factors experts met with researchers at MITRE CAASD for a broad review of programs involving human factors. The teams discussed potential areas for collaboration and established specific contacts between organizations and researchers with intersecting interests. MITRE CAASD representatives attended a human factors lab research program review at the FAA Technical Center Human Factors laboratory soon after the TIM to discuss potential collaboration on air traffic control display automation.

FAA and NASA researchers likewise discussed areas for potential collaboration, including FAA participation on NASA's Human Performance and Modeling advisory committee. FAA researchers attended a NASA-sponsored intra-agency human factors symposium to highlight important NASA research.

Findings

The TIMs have been useful as a forum for getting to know members of the human factors research community, learning about the research that each organization is conducting, and sharing information about challenges researchers have faced.

Even more importantly, the intent of collaboration and information sharing is gaining momentum as a business imperative because of dwindling R&D budgets. In the present FAA environment it is imperative that the human factors community do a better job at cost-benefit analysis as well as collaborating earlier in the research cycle to ensure an understanding of full ownership costs. The human factors research community understands this and has agreed that collaboration should be continuous and worked at project levels. At the same time, strict adherence to requirements-driven research will hinder innovation and creativity. Organizations need to maintain a balance between requirements and innovation. Sharing labs and other unique facilities, including NAS simulation capabilities, would yield significant cost savings and encourage collaboration.

The lessons learned from prior R&D experiences challenge the research community to manage expectations not only within the R&D community but with the customers of R&D capabilities. Lessons learned further challenge organizations to delineate roles and responsibilities including a more balanced FAA/NASA approach; define research processes and decision points in research activities to determine what capabilities progress to field implementation and what capabilities do not; and systematically audit and inventory current and required laboratories and facilities, personnel resources, and research capabilities.

Conclusions

TIM participants agree that the transition from research to operational prototypes to development and fielding is complex and replete with human factors challenges. Moreover, they agree that FAA needs to establish an integrated and coordinated human factors R&D program that focuses on user needs, avoids duplication of effort, and leverages all research capabilities including people and laboratories. It is important to be clear about roles and responsibilities between researchers and their organizations. Fortunately, ongoing collaboration, as evidenced by post-TIM activities, is gaining momentum. Collaboration, not competition, is critical to success.

References

DellaRocco, P., et.al. (2002) *Collocation of User Request Evaluation Tool, Traffic Management Advisor, and Controller Pilot Data Link Communications: An Initial Human Factors Evaluation*, FAA Technical Center Report.

Federal Aviation Administration and National Aeronautics and Space Administration (2003c). *Integrated Plan for Air Traffic Management Research and Technology Development, Version 7.0, Volume I: Plan Overview and Joint Research Project Descriptions.* Washington, DC: U.S. Department of Transportation.

Federal Aviation Administration, *Memorandum to the Vice President for Operations Planning on the Transition of R&D into NAS Implementation*, March 22, 2004, Washington, DC.

Krois, P., Mogford, R., & Rehmann, J. *Integrating Human Factors Issues in Technology Readiness Levels for Air Traffic Control Research.* Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting, 2003.

EXTERNAL AIDS AND AGE DIFFERENCES IN PILOT COMMUNICATION

Daniel Morrow, **Dervon Chang**, **Christopher Wickens**, **Esa Rantane**n, and **Liza Raquel**
University of Illinois at Urbana-Champaign

The use of external aids (e.g., a kneepad) can reduce the demands of Air Traffic Control (ATC) communication on pilots' working memory during routine flight. Older pilots may especially benefit from such aids because of age-related declines in working memory, although cognitive declines may impair the ability to coordinate the use of these aids with concurrent flight navigation and control tasks. We investigated the use of two external aids that may vary in ease of coordination: a conventional knee pad and an electronic notepad, or e-pad. Participants were 6 older (50-65 years) and 6 younger (20-40) active instrument-rated pilots. While in a Frasca flight simulator, they listened to and read back complex (four-instruction) ATC messages while using the kneepad, e-pad, or no aid. Readback accuracy was analyzed by an Age x Aid x Instruction type (Heading, Altitude, Speed) ANOVA with Aid and Instruction as repeated measures. Accuracy was higher when pilots used either aid compared to no aid, and lower for older pilots. The findings suggested a greater aid benefit for older pilots, with a smaller age difference in the two aid conditions than in the no-aid condition. While the Age x Aid interaction was not significant, this interaction was significant for the altitude instruction readbacks. Despite the small sample size, our study replicates note-taking (kneepad) benefits for older as well as younger pilots' communication, and extends these findings to the novel e-pad. Results of a usability survey helped improve the e-pad interface. We will next investigate potential attentional costs of these aids for task coordination during simulated Frasca flight, as well as their benefits for communication.

Introduction

Communication in complex environments such as piloting and driving places heavy demands on operators' cognitive resources, occasionally contributing to problems that reduce safety and efficiency. It may especially challenge older pilots who tend to experience declines in working memory. External aids such as note-taking may help older pilots manage these demands, especially if these aids are part of the pilots' skill repertoire.

Note-taking provides environmental support (Craik & Jennings, 1992) that reduces working memory constraints on responding to Air Traffic Control (ATC) messages. Morrow, Ridolfo et al. (2003) found that note-taking reduced age differences among pilots on a readback task compared to a no-aid condition. However, note-taking in that study was investigated in a communication-only rather than multi-task environment typical of piloting. Note-taking involves visual components, and thus, according to multiple resource theory, may compete with concurrent visual tasks such as flight control for modality-specific attentional resources (Helleberg & Wickens, 2003). For example, writing on a kneepad often incurs heads-down time, drawing attention from the instrument panel, which supports flight control. Heads-down time can greatly affect a pilot's situation awareness (SA), especially during critical out-of-the-window times (e.g., detecting traffic) (Endsley, & Garland, 2000). Thus, external aids should be designed to minimize visual competition with concurrent flying tasks at hand, as well as to support communication.

Such high demands on cognitive resources may especially challenge older pilots because of their declining ability to allocate resources to multiple tasks. Tsang and Shaner (1998) found that older pilots exhibited age-related declines in time-sharing tasks under high levels of attentional demand. Time-sharing in this case related to performing concurrent tasks that were similar to navigating the plane and listening to ATC communications.

On the other hand, older pilots' high levels of expertise may help them compensate for these age-related cognitive declines, Studies of expertise in pilot communication and decision-making have found greater benefits for higher levels of expertise (e.g., Wickens, Stokes, Barnett, & Hyman, 1993; Wiggins & O'Hare, 1995). Morrow, Ridolfo et al. (2003) found that note-taking eliminated age differences in readback accuracy among pilots but not nonpilots. There is also some evidence that expertise reduces age differences in the ability to perform multiple tasks (Lassiter et al., 1997; Tsang & Shaner, 1998). Even so, expertise may be less likely to eliminate age-related declines in communication in complex, multi-task environments, such as aviation. Therefore, we investigated external aids that may vary in their ease and effectiveness of use in single- and multi-task flying environments.

We compared conventional note-taking (*kneepad*) with an electronic notepad positioned adjacent to the instrument panel *(e-pad)*. The e-pad resembles Mode Control Panel interfaces common in commercial flight management systems, but it functioned only as an external aid in the present study. All participants were General Aviation, and were not familiar with this type of interface. Although the kneepad is more familiar to pilots, it may be easier to coordinate the e-pad with concurrent tasks because it is more integrated with the flight instruments, reducing heads-down time. Both aids should reduce age differences in communication compared to a no-aid condition (see Morrow, Ridolfo et al. 2003), and the e-pad is more likely to reduce age differences as concurrent task demands increase. Because use of external aids depends on the costs associated with perceptual access of information from the aid compared to accessing the information from memory (Fu & Gray, 2000), we first conducted the present study to explore the usability of the two aids primarily in a single-task environment.

Method

Participants

Twelve instrumented-rated pilots participated (minimum 500 total flight hours). Six were older (50-64 years), and six younger (20-40 years).

Table 1. *Mean Demographic and Cognitive Ability Scores*

	Older N=5[1]	Younger N=6	*Mean*	*Age t(10)*
Age	53.8	25.5	*38.36*	
Educ (years)	17.4	15.8	*16.55*	*1.0*
Speed_letter[2]	10.4	12.8	*11.7*	*2.1**
Speed_pattern[2]	18.7	19.75	*19.27*	*<1.0*
Total Flight hours	2975.6	1342.7	*2084.9*	*1.2*
Hours last 12 months	49.9	139.4	*156.6*	*1.7*
Total IFR hours	488.25	139.38	*278.9*	*2.1*
Self-rated Health	5.5	6.3	*6.0*	*1.0*

* $\underline{p}<.05$

1. Only 5 of the 6 older pilots who participated filled out demographic and pilot questionnaires
2. Letter and Pattern Comparison tasks, a measure of processing speed (Salthouse & Babcock, 1991)

The two age groups did not differ significantly in years of education, flight experience, or self-rated health (see Table 1). We also included a measure often used to index speed of mental processing (Letter and Pattern Comparison tasks, Salthouse & Babcock, 1991). Typical of cognitive aging studies, younger pilots outscored older pilots on the Letter Comparison measure (the difference was in the same direction, but nonsignificant for the Pattern measure).

Apparatus

Participants performed all ATC communication tasks in a Frasca 142 flight simulator, configured as a single-engine, fixed wing light aircraft, including a full set of flight displays on the instrument panel and radio, and a three-screen out-the-window display. A touch screen display served as the e-pad (see Figure 1) and was placed adjacent to the instrument panel (Figure 2).

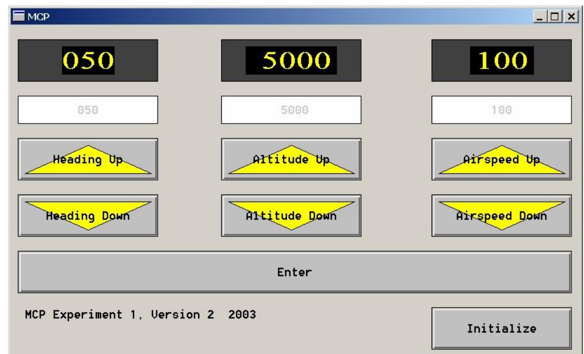


Figure 1. *E-pad touch screen display*

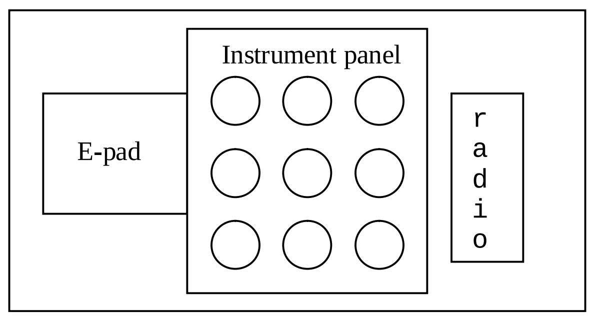


Figure 2. *E-pad, instrument, and radio layout*

Procedure

Participants listened to pre-recorded ATC messages for four flight scenarios. Each message directed the pilot to make heading, altitude, speed, and frequency or squawk changes (i.e., 4 instructions). In three scenarios, the participants used the kneepad, e-pad, or no aid while only listening to and reading back each message. In the fourth scenario, they flew the described route in the simulator as well as using the e-pad to support communication. They were given a practice session using the e-pad to familiarize themselves with this novel touch screen display.

In the kneepad condition, participants listened to the ATC instructions and wrote any notes on a kneepad strapped to their leg. In the e-pad condition, they entered heading, altitude, and speed changes into the touch screen display (see Figure 1). Notes and e-pad button press responses were scored for comparison to readback accuracy. No aid was available in the third condition, so pilots read back messages from memory. In all three conditions, participants were allowed to use the radio, located to the right of the instrument panel, to enter frequencies or squawks (Figure 2). Participants could also ask for ATC message repeats. Readbacks and requests for repeats were tape recorded for later scoring and analysis.

After completing the first three scenarios (communication-only conditions), participants completed a questionnaire about the ease of using the e-pad display and task workload, including comparisons of the e-pad and kneepad aids. The same questionnaire was given after the fourth scenario in order to investigate whether e-pad usability varied in single task (communication only) and multi-task (i.e., communicating and flying) environments. At the end of the session, all participants completed a demographics and pilot experience questionnaire and the Letter and Pattern Comparison tasks (Salthouse & Babcock, 1991).

Results

Readback Accuracy

Readback accuracy (mean percent correct instructions repeated) was analyzed by an Age x Aid (kneepad, e-pad, no aid) x Instruction type (heading, altitude, speed) ANOVA with Aid and Instruction as repeated measures. As shown in Figure 3, accuracy was higher when pilots used either aid compared to no aid, $\underline{F}(2,20)=31.8$, $\underline{p} < .001$, and slightly lower for older pilots, $\underline{F}(1,10)=5.6$, $\underline{p} < .05$. There was also an effect of instruction, (H=97%, A=94%, S=91% correct, $\underline{F}(2,20)=6.5$, $\underline{p} < .01$, which is difficult to interpret because the three instruction typess were always presented in the same (standard) message positions (heading first, speed last).

While the Age x Aid interaction was not significant, $\underline{F}(2,20)=1.1$, the pattern in Figure 3 suggests a greater aid benefit for older pilots, with a smaller age difference in the two aid conditions (Y=100%, O=98%) than in the no-aid condition (Y=88%, O=81%). Analysis of age and aid effects for each instruction revealed an aid benefit for all three instructions, but only a significant age decline for the altitudes ($\underline{p} < .01$; $\underline{p} > .10$ for heading and speed instructions). Moreover, the Aid x Age interaction was significant for altitudes, the most age-sensitive readback measure $\underline{F}(2,20)=4.1$, $\underline{p} < .05$.

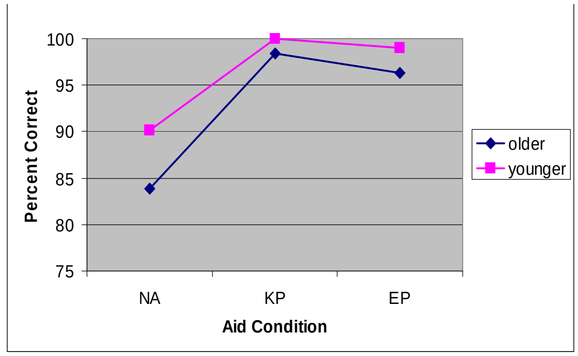


Figure 3. *Readback Accuracy.*

All aid values (i.e., notes written on the kneepad and values entered into the e-pad display) were also scored. There were no discrepancies between the accuracy of aids and readbacks.

Requests for Message Repeat

Analysis of mean number of requests for ATC message repeats revealed a similar effect of aid (KP=0.0, EP=0.48, NA=2.0 mean requests), $\underline{F}(2,18)=5.9$, $\underline{p} < .05$. The age difference in requests was not significant.

E-pad Benefits in Single- and Multi-task Environments

Mean readback accuracy in the two task conditions was analyzed by an Age x Task (single-task, multi-task) ANOVA with the latter a repeated measure. Performance did not vary by task condition (Single: 99%, Multiple: 98%), $\underline{F} < 1.0$, or by age group (Y= 99%, O=97% $\underline{F}(1,9)=2.3$, $\underline{p} > .10$. While null findings must be treated cautiously because of the small sample size, this analysis suggests that communication benefits from the e-pad were not reduced by performing multiple tasks for older as well as younger pilots.

Discussion

Older and younger pilots more accurately read back complex ATC messages when using either the kneepad or the e-pad, compared to no aid. There was also some evidence that both aids reduced age differences in communication accuracy, consistent with environmental support theory (Craik & Jennings, 1992). This finding replicates the earlier finding of note-taking benefits for older as well as younger pilots' communication (Morrow et al., 2003), and extends these findings to the novel e-pad aid.

Both age groups were also more likely to request repeats of the ATC messages in the no aid condition. While either aid provided an external form of working memory for the readback task, pilots required more exposure to the information (i.e., more message presentations) when relying on memory, and they still made more errors without the support of the aids. Moreover, in actual operations, the increased frequency of ask for and receiving clarification from ATC would decrease communication efficiency and potentially impair concurrent task performance in multi-task environments. The absence of age differences in requests for repeat may reflect age-related differences in communication style that mask age differences in memory, or the possibility that the present study did not impose sufficient task demands to produce age differences on this measure. A follow-up study (see below) will vary task difficulty in multi-task environments to examine the latter possibility.

Limitations of the Present Study

The small sample size limits our ability to identify the effects of pilot age and external aids on communication. Nonetheless, the age difference on the Letter Comparison measure of processing speed suggests that the pilots in our sample were experiencing typical age-related changes in cognitive abilities. In addition, the pattern of aid benefits for the older and younger groups (smaller age differences in communication accuracy for the aid versus no-aid conditions) is similar to earlier studies with larger sample sizes (Morrow et al., 2003).

Designing the E-pad: Usability Issues

The fact that older pilots tend to experience typical age-related changes in speed of processing and working memory, coupled with findings that operators are less likely to use external aids as the cost of perceptual access increases (e.g., Fu & Gray, 2000), has important implications for designing novel aids such as the e-pad. To the extent that using the aid exacts perceptual-motor costs, older pilots may be less willing to use them. Therefore, an important goal of the present study was to improve e-pad usability. Questionnaire findings suggested that both age groups actually preferred using the kneepad over the e-pad. Participants' comments suggested the importance of the kneepad's familiarity. Although participants were given practice with the e-pad, the amount of practice could not compare pilots' years of experience with the kneepad. Consistent with this, workload ratings suggested greater difficulty using the e-pad in the multi-task condition (communication and flight control). Typically, (right-handed) pilots use their left hand on the yoke while writing on the kneepad with their right hand, as well as using their right hand to input radio frequencies and squawks. With the e-pad positioned to the left of the flight instruments and controls, pilots pointed out that they would either have to use their left hand to input into the e-pad and then switch to the right hand to input into the radio or use their right hand to cross over the yoke to input into the e-pad. Neither felt natural to them.

Other comments included the use of some unnecessary displays and controls (*enter* button), and lack of haptic feedback when pressing buttons. In response to these concerns, the e-pad interface was modified for the primary study. Changes included eliminating extraneous displays and buttons and reducing the screen size so that the display could be moved closer to the flight instruments in order to be more integrated with the instruments. The new display reduces clutter without reducing button size, and decreases screen brightness that interfered with lighting for the instrument panel (see Figure 4).

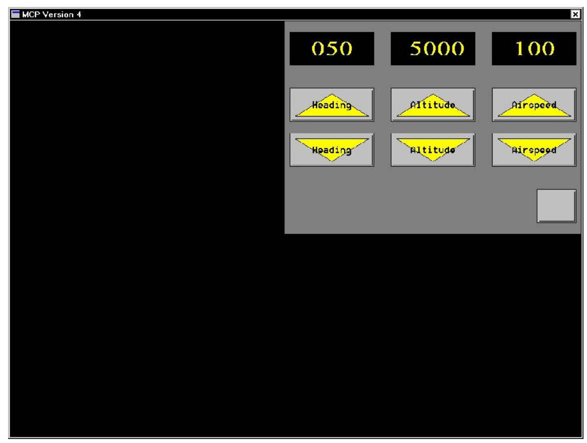


Figure 4. *New E-pad display screen*

Next Steps

In a follow-up study, we are now investigating whether the e-pad is more effective than the kneepad in reducing age differences in communication performance under demanding multi-task (navigation and flight control as well as communication tasks). In addition to the communication-only conditions used in the present study, scenarios are included that require the pilots to fly the route described by the messages while looking out the window for traffic as well as communicating with ATC. Flight performance and eye-tracking measures will be used to assess the impact of the external aids on communication performance and attentional requirements of coordinating these aids with the concurrent flight tasks

Acknowledgments

We thank Maggie Rudolphi for developing the scenarios, Don Talleur for expert advice on the simulation and experimental design, and Jonathan Sivier and Roger Marsh for developing the e-pad. Support provided by NIA grant PHS 1 P30 AG023101. Address correspondence to Daniel Morrow, University of Illinois Institute of Aviation, Aviation Human Factors Division, Willard Airport, #1 Airport Rd, Savoy IL, 244-8757 .

References

Craik, F. I. M., & Jennings, J. M. (1992). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), The handbook of aging and cognition (pp. 51-110). Hillsdale, NJ: Erlbaum.

Endsley, M. R. & Garland, D. J. (2000). Situation awareness analysis and measurement. Mahwah, NJ: Lawrence Erlbaum Associates.

Fu, W., & Gray, W. D. (2000). Memory versus perceptual-motor tradeoffs in a blocks world task. Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society (pp. 154-159). Hillsdale, NJ: Erlbaum.. Hillsdale, NJ: Erlbaum.

Helleberg, J. R., & Wickens, C. D. (2003). Effects of data link modality and display redundancy on pilot performance: An attentional perspective. International Journal of Aviation Psychology, 13, 189-210.

Lassiter, D., Morrow, D.G., Hinson, G., Miller, M., & Hambrick, D. (1997). Expertise and age effects on pilot mental workload in a simulated aviation task. in W.A. Rogers (Ed.), Designing for an aging population: Ten years of human factors/ergonomics research (pp 226-230). Santa Monica, CA: Human Factors & Ergonomics Society.

Morrow, D.G., Ridolfo, H.E., Menard, W.E., Sanborn, A., Stine-Morrow, E.A.L. et al. (2003). Environmental support promotes expertise-based mitigation of age differences in pilot communication tasks. Psychology and Aging, 18, 268-284.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. Developmental Psychology, 27, 763-776

Tsang, P.S., & Shaner, T.L. (1998). Age, attention, expertise, and time-sharing performance. Psychology and Aging, 13, 323-347.

Wickens, C. D., Stokes, A., Barnett, B., & Hyman, F. (1993). The effects of stress on pilot judgment in a MIDIS simulator. In O. Svenson & A. J. Maule (Eds.), Time pressure and stress in human judgment and decision-making (pp. 271-292). New York: Plenum Press.

Wiggins, M. & O'Hare, D. (1995). Expertise in aeronautical weather-related decision-making: A cross-sectional analysis of general aviation pilots. Journal of Experimental psychology: Applied. 1, 305-320.

THE RECORDING OF FLIGHT TIME BY PILOTS

Rudolf G. Mortimer, PhD
Urbana, IL

Flight time is an important index of the flight exposure of pilots and is used in estimating the accident rates of pilots, but we are not aware of prior attempts to determine how the underlying data are obtained. This study used two surveys of civilian U.S. pilots to learn how they measured and recorded flight time. Pilots used a timer, watch, the Hobbs meter or tachometer to measure the duration of flights. Professional pilots flying on scheduled flights used company or block time as the criterion of the duration of a flight, while 79% of private pilots used the time from engine(s)-on to shut-down and others the time from take-off to landing. About 80% of the pilots made a temporary or permanent record of the hours flown on the day of the flight. But, only about 71% of pilots with class III medical certificates, 58% with class II and 42% with class I reviewed their flight logs before the medical exam. The accuracy of the flight hours data could be improved if pilots used a uniform criterion of flight time, recorded it before leaving the cockpit and checked their flight logs before the medical exam.

Introduction

The objectives of this study were to learn the criteria used by pilots in determining the duration of a flight and how pilots went about recording their flying hours. In addition, the habits of pilots of different medical classes, age and flight experience may have a bearing on those activities and also on whether or not the pilots reviewed their logs of flight time before the medical exam. The medical exam is relevant to this issue of flight exposure because it is at that time that pilots are asked to complete a medical history form, which includes a section in which they indicate the number of hours they have flown in the last six months and their total flying hours. The data of flight time are important because they have been used by researchers as the measure of flight exposure of pilots and, hence, evaluations of the risks associated with various characteristics of pilots are dependent on the quality of the exposure measures.

For example, some studies have been done to look at the role of the age of pilots on accidents (Golaszewski, 1983; Mortimer,1991; Kay et al., 1993). Those studies used the FAA's Medical History File to compute the pilots' flight exposure and, hence, the validity of the results is contingent upon the accuracy of the flight time that pilots reported during their medical exam. This raised a number of questions such as, how and when do pilots record their hours flown and what proportion of pilots usually refer to their logbooks before the medical? This study has looked at these issues.

Method

We conducted two separate mail surveys of civilian pilots, resident in all the FAA's geographic regions in the United States. The survey was mailed to a non representative sample of 1413 pilots 40% of whom held a class I medical certificate and 30% each a class II or III. Also, in each group, half the pilots had single-engine and half multi-engine ratings. We also gave 50 questionnaires to pilots who attended an aviation safety seminar to mail back to us, in order to increase the number of pilots with class II and III certificates in our sample, who had a low mail response rate. The usable survey response rate was 36%. The sample was specifically not intended to be representative by medical certificate of the U.S. pilot population but was intended to ensure a sufficient number of cases in each medical class/ engine rating category to make estimates about their behavior in measuring and recording their flight hours.

Characteristics of the Sample

The youngest pilot in our sample was 18 and the oldest was 82 years old. 43% of the respondents had a Class I medical certificate, 30% Class II, 27% Class III. Compared to the age and class composition of the civilian pilot population at the national level (US. DOT, 1991), younger and Class III pilots were under-represented in our sample, while pilots between 30 and 59 years of age and Class I pilots were over-represented. The hours reported by the pilots in our sample are somewhat greater than those reported at the national level. In addition, 3.2% of the respondents were female.

Results

How Flight Time is Measured

The tools that pilots used to measure the time of a flight differed according to their medical class ($P<0.01$, Chi-Square). Class III pilots clearly favored (59%) the Hobbs meter, 23% used the elapsed time on the tachometer and 17% a watch or timer. Class II pilots used the Hobbs meter (39%) and a watch or timer (33%) almost equally, 26% used the tachometer and 1.5% company scheduled time. Class I pilots mostly (46%) used a watch or timer, 27% the Hobbs meter, 26% company schedule and only 2% a tachometer (Table 1).

The Criterion of Flight Time

The differences in the criteria used as flight time by the pilots were significantly ($P<0.01$) associated with their medical classes (Table 2). Just over half of pilots with a Class I medical certificate used company scheduled or block time while 86% of those with a Class II medical and 97% of those with a Class III medical used the time from starting to stopping the engines or the time from take-off to landing. Also, Class II pilots used the time from take-off to landing more often than those with Class I or III certificates (30% v. 19%, 17%).

When Pilots Make a Temporary or Permanent Record of Flight Time

Slightly over half of the pilots (55%) made a temporary or permanent record of the duration of a flight before leaving the cockpit after a flight and another 25% later on the same day (Table 3). Thus, a total of 80% of the pilots made a record of their flight time some time on the day of the flight. Another 7% of pilots recorded later than the day of the flight, but before the next flight. There were significant ($P<0.01$) differences when flight time was recorded. Only 47% of Class II pilots recorded flight time before leaving the cockpit compared with 58% and 60% of Class I and III, respectively. Class I pilots made a record of flight time later on the day of the flight less often (18%) than Class II and III (32% and 29%) respectively. Overall, 89% of Class III pilots made a record of their flight time some time on the day of the flight compared with 76% of class I and 79% of Class II pilots.

Pilots Who Do Not Record Flight Time

A few pilots (4%) did not record flight time. They were mostly professional pilots with an Airline transport pilot certificate and their hours were recorded by their airline. The remainder, estimated at less than 1%, may not record their hours.

Logbook Referral Prior to the Medical

About 54% of the pilots usually referred to the logbook before the medical exam, and the percentage of pilots who did so varied by their class of medical certificate and flight experience. 41% with class I medical certificates, 57% with class II and 71% with class III referred to their logbooks before their medical exam. There was not a significant association between the age of the pilots and their referral to their flight logs before the medical.

Discussion

This study has provided information about the way pilots recorded the hours flown. Since this is the most frequently used measure of the flight exposure of pilots, the manner in which it is obtained is important because it forms the basis of risk assessments of various characteristics of pilots. However, the criteria that pilots use and if and when they record the time of flight and if they review their logs, before responding to the question on the medical history form asking their flight experience, have not been studied previously.

How Flight Time Was Recorded

This study has shown that various methods are used by pilots to measure the time of a flight, but most general aviation pilots rely on the Hobbs meter or the tachometer or a watch or timer. In this sample, 77% of class III (private) pilots used the Hobbs or a watch or timer. The tachometer was used by the rest. The same (73%) was basically the case for pilots with class II medical certificates. Company schedule was used by 26% of class I pilots and 73% used the Hobbs meter, watch or timer. It is important to remember that the stratified sampling procedure that was used would include non-professional, single engine rated pilots among up to half of those with class I medical certificates.

Criteria

There were also various criteria used to denote the time of a flight which differed by medical class. Company schedule or block in/out time was used by about half of class I pilots. Those would be professional airline pilots. Class II and III pilots mostly used the time from engine(s)-on to shut-down. The time from take-off to landing was used by 30% of those with class II medicals and by about 20% of those with class I and III certificates. While engine-on to shut-down will maximize the time, by including the time on the ground and taxi time, using the take-off to landing will minimize it. On short flights, of an hour or so, the ground portion can be 25% or more. That suggests that the hours of many general aviation pilots may be inflated compared with the official definition. The official definition of flight time (FAR 14 CFR, Ch 1) is the time from the moment that the aircraft first moves until it stops after the next landing (block to block time). Other than block-to-block time, none of the criteria used by the pilots in our sample met the FAA's criterion.

When Flight Time was Recorded.

Between 47% and 59% of pilots recorded the flight time before leaving the cockpit and about 80% at some time on the day of the flight. Class I and III pilots made a record before leaving the cockpit more often than class II pilots, so memory was less of a factor for the former. Only one class III pilot out of 135 kept no record in this sample.

Reviewing the Log Before the Medical

About 54% of the sample reviewed their logs of flight time before the medical exam. Age alone was not a factor in the frequency of referral to the log before the medical, but medical class and total flight hours were significant factors affecting this behavior.

Pilots have to report the hours they flew in the prior six months and their total hours on the form they complete when applying to renew their FAA medical certificate. This form is completed immediately before the exam itself.

The results show that class III pilots refer to their logs before the medical substantially more often than class I or II. Overall, 71% of class III, 57% of class II and 41% of class I pilots checked their logs before the medical, and pilots aged less than 30 or 60 and above were more likely to do so as well as those with less than 4000 hours.

Sources of Errors

The study has shown that errors in the flight hours reported by pilots can occur because of various factors: 1. Pilots use various means by which to measure flight time (e.g., Hobbs, timer, company schedule); 2. Pilots use various criteria (e.g., take-off to landing, engine(s)-on to shut-down, block-block); 3. Some pilots fail to record the index of flight time before leaving the cockpit, which necessitates later reliance on memory; and 4. Some pilots do not refer to their logs before completing the medical history form, basing their response on memory.

The errors introduced by the method of measurement used is probably small in affecting the time of a flight, but may have a bearing on the pilots' ability to recall the value if it is not immediately recorded. A direct reading of the time of a flight, by clock or timer, should be easier to memorize correctly than a Hobbs or tachometer reading, which only show the cumulative time --not the time of a flight.

The criteria of the duration of a flight will converge if the flights are relatively long. But, in short flights, say of an hour, there can be substantial differences between the time from engine(s)-on to-off and the time from take-off to landing.

The hours recorded by pilots in their logs can be expected to be least affected by memory for class I and III pilots because more of them made a temporary or permanent record before leaving the cockpit. However, 45% of class I pilots used a watch or timer as did 33% of class II and only 17% of class III pilots to measure flight time (Table 1). If it is true that pilots can recall the duration of a flight better from a watch or timer, which reads the actual flight time in hours and minutes, than the cumulative time on a Hobbs or tachometer, the accuracy of recordings made after leaving the cockpit, taking account of inaccuracies due to memory, should be better for class I and II pilots than class III.

Considering also that another 26% of class I pilots (Table 1) used company or block time, suggests that the hours recorded by or for class I pilots may be among the most accurate, followed by those for class III pilots because of their greatest tendency to record flight time before leaving the cockpit.

Finally, the accuracy of the hours reported on the medical history form should be a function of whether the pilots referred to their logs beforehand. We have already seen that class III pilots check their logs (71%) much more than class II (57%) and class I (41%). For the latter groups in particular, the ability to recall flight time will be a factor. That ability will be affected by how often pilots update their logs and their consistency in the hours they fly.

The medical history form requests information of total time and also time in the prior six months. Pilots who frequently update their logs can be expected to estimate their total hours quite accurately. However, their ability to estimate the hours flown in a prior period of six months may be much worse. That is because logbooks show the cumulative hours and a special effort has to be made to add the hours over any six month or other period.

Therefore, pilots who do not refer to their logs prior to the medical may be expected to estimate their hours in the prior six months with greater relative error than those who do and with greater relative error than their total time.

However, professional pilots fly a relatively fixed number of hours per month. Even if they do not refer to their logs before the medical, they should be able to estimate their hours quite well. Therefore, even though they refer to their logs least often, class I pilots as a group, may report the most veridical hours. Class II pilots include some professional pilots who also fly on a regular basis and they referred to their logs more than class I pilots but less than class III pilots. Table 3 shows that they made a record before leaving the cockpit least often so that their records are most subject to errors of memory. It is hypothesized that their reports on the medical form may be the least veridical.

Class III pilots recorded flight time before leaving the cockpit or later on the day of the flight and referred to their logs before the medical most often, which will enhance the accuracy of their reports. The 40% of them who did not record flight time before leaving the cockpit did have to remember the reading on the Hobbs or tachometer, which may lead to errors.

Recommendations

The study has shown that the procedures used by pilots on which they base their reports of flight hours are quite variable and subject to intrinsic errors. At least three of the major sources of error could be removed if pilots could be induced to (1) record flight time before leaving the cockpit and (2) check their logs or company records before the medical and (3) use the same criterion of flight time.

While flight time is a basic measure of exposure of pilots, we recommend that more extensive studies be done, not only to verify our findings, but to extend our knowledge of the types of conditions to which pilots are exposed. The risks associated with a broader range of factors affecting flight safety could then be ascertained.

References

Aeromedical Certification Statistical Handbook (1991). Civil Aeromedical Institute, U.S. Department of Transportation.

Golaszewski, R. (1983). The influence of total flight time, recent flight time and age on pilot accident rates. Final Report, U.S. DOT., Transportation Systems Center and FAA Safety Analysis Division.

Kay, E.D., Harris, R.M., Voros, R.S., Hillman, D.J., Hyland, D.T. Deimer, J.D. (1993). Age 60 project: Consolidated data base experiments, Final Report, Contract DTFA-02-90-90125, CAMI-FAA, Hilton Systems.

Mortimer, R.G. (1991). Some factors Associated with pilot age in General Aviation crashes. Proceedings, Association of Aviation Psychologists, Columbus, Ohio, 770-775.

Table 1. *How Flight Time was Measured by Medical Class, in Percent*

How Measured	Class I	Class II	Class III
Hobbs Meter	27	39	59
Tachometer	2	26	23
Watch/Timer	46	33	17
Company Schedule	26	2	2

Table 2. *The Criteria used as Flight Time by Medical Class, in Percent*

Criterion	Class I	Class II	Class III	%
Engine(s) on to shut down	30	65	80	53
Take-off to landing	19	30	17	22
Company schedule/ Block in/out	51	5	3	25

Table 3. *Temporary or Permanent Recording of Flight Time by Medical Class, in Percent*

Time	Class I	Class II	Class III	%
Before leaving cockpit	58	47	60	55
Later on the day of the flight	18	32	29	25
Other	18	17	10	15
Do not record flight time	6	4	2	4

AN EVALUATION OF SCANNING OF INTEGRATED HAZARD DISPLAYS AS A FUNCTION OF SIZE AND EVENT DETECTION PERFORMANCE

Emily K. Muthard
Christopher D. Wickens
University of Illinois
Institute of Illinois, Human Factors Division
Savoy, Illinois

The present study was designed to assess the influence of display enlargement on pilot scanning patterns and event detection performance. Nineteen pilots monitored an integrated hazard display for changes in the altitude or heading of traffic aircraft and weather systems. Analyses revealed that event detection accuracy and response time were unaffected by display size, suggesting pilots compensated for display enlargements by strategically widening scanning patterns. While eye movement data revealed that attention was allocated to the peripheral display regions regardless of display size, individuals who were poor at detecting events were less likely to attend to these display areas. The results suggest that the attention allocation patterns of pilots are adaptive and flexible and that such flexibility leads to higher performance in attentional tasks.

Introduction

With continuing advances in the technological capabilities of aviation displays, designers are now able to implement large scale displays that portray a top-down view of the environmental hazards that exist within a large region of the airspace. The integrated hazard display represents one example of these advanced displays.

The integrated hazard display depicts traffic, terrain, and weather hazards in a single, unified panel. This integrated display layout allows the pilot to easily monitor the dynamic airspace for changes in the lateral and vertical behavior of traffic aircraft and weather systems. This monitoring task requires that the pilot identify changes in the movement or location of a hazard from one moment to the next. Unfortunately, observers have been shown to be poor at change detection, reflecting instead a tendency toward “change blindness” (Carpenter, 2001; Rensink, 2002; Simons, 2000). Change blindness refers to the inability that observers have in detecting changes that occur beyond the focus of attention (Levin & Simons, 1997; Muthard & Wickens, 2002; Pringle, Irwin, Kramer, & Atchley, 2001; Rensink, O’Regan, & Clark), particularly as events occur at locations that are increasingly more distant from the fovea (Pringle et al., 2001).

Any enlargement to a display augments the area that must be scanned and increases the proportion of the display that is located within the observer’s periphery. Thus, it reasons that such changes to display size may also hinder change detection, as more effort must be used to access peripherally located information. Wickens (1992) proposes a model of information access effort that describes this relationship. In this model, the effort required to access information is proposed to increase nonlinearly with eye movements and head movements of increasing magnitudes. Consequently, small displays, which can be monitored effortlessly with short saccades, do not induce large scanning costs. Conversely, when larger displays exceed twenty degrees of visual angle, observers must begin to use head movements to access peripheral information (Bahill, Adler, & Stark, 1975), which become increasingly effortful as they increase in magnitude (Wickens, 1992).

Two models have been proposed to examine the potential surveillance strategies that pilots might employ in response to display enlargement. The first model is supported by the presented research on attentional effort and is termed the *effort conservation* model. Under this model, pilots do not invest the effort needed to access the most peripherally located information. As a result, detection performance for the most peripheral changes suffers proportionally more with display enlargements. In fact, under an extreme effort conservation model, the ratio of performance decrement with display enlargement should be equal to the ratio of the sizes of the displays. The second model of *strategic compensation* postulates that the pilot realizes that peripheral events will go unnoticed if the display perimeter is not monitored. As a result, the pilot strategically adapts and enlarges his scanning area, despite the extra resources that must be deployed to do so. For pilots strategically compensating for display size, surveillance performance would not differ across display sizes.

While understanding the potential hindrance that display enlargement may pose to attention-based tasks is important, only a handful of studies have explicitly examined this relationship. Enoch (1959) asked participants to search for a Landolt C, which was presented on aerial maps that ranged in size from 3 to 51 degrees of visual angle. Enoch's work indicated that display enlargement resulted in shorter fixations and longer saccades. Fixation length was reduced because fixations to the most peripherally located display regions were difficult to maintain for extended time periods and because additional search time was needed to make longer saccades to reach these regions. Enoch (1959) also reported that the concentration of fixations in search was located in the center of the maps, particularly for displays of larger size. Kroft and Wickens (2003) also examined search for hazards on sectional charts. While Enoch (1959) reported disadvantage to larger displays, Kroft and Wickens (2003) determined that search was inhibited by small displays, largely because of the reduced legibility of symbols and text. While this pair of studies provides some indication of the influence that display size may have on performance, both examine the task of goal-directed search rather than surveillance.

The present study seeks to examine the influence of display enlargement on the task of hazard surveillance within the context of the proposed effort conservation and strategic compensation models. In a low fidelity simulation, pilots were asked to monitor an integrated hazard display for changes in the altitude, airspeed, and trajectory of traffic aircraft and weather systems, while also flying the aircraft. Change detection performance was assessed as a function of event location and display size. Eye movement data was also collected as a measure of surveillance. To the extent that pilots employed an effort conservation strategy, change detection performance should be reduced with display enlargements, particularly for the most peripheral changes. Scanning to the display perimeter should also be reduced. If pilots were able to strategically compensate for display enlargement by widening scanning patterns, however, the proportion of fixations in the outer display regions and change detection performance should not be affected by size.

Methods

Subjects

Nineteen pilots from the University of Illinois, Institute of Aviation participated in the study. These pilots ranged in age from 19 to 23 years (M = 21 years) and all were male. Participants had an average of 226 flight hours of experience. Six pilots had private licenses while the remaining thirteen were instrument certified.

Display

Pilots were shown an integrated hazard display that depicted traffic aircraft and weather systems overlain on a topographical map, as shown in Figure 1. The topographical map was based on the National Oceanic and Atmospheric Administrations (NOAA) sectional aeronautical chart. Traffic aircraft were depicted with small aircraft icons and digital data tags that included the aircraft's call sign, altitude, heading, and airspeed. Weather systems were portrayed as a series of concentric circles. The altitude of weather tops were shown with data tags located in the center of each weather system.

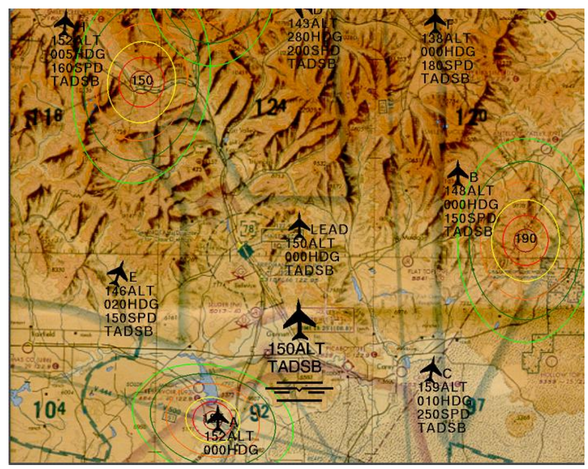


Figure 1. *Integrated hazard display. Ownship is located in the center of the display.*

Ownship was depicted with a large aircraft icon and was always located in the center of the display. Ownship remained stationary at this location and traffic aircraft and weather moved relative to ownship. An attitude directional indicator, which depicted only pitch, was located directly below ownship to assist in altitude control.

The integrated hazard display was presented in three sizes. The small display measured 8.9 by 6.4 cm and encompassed 10 by 7 degrees of visual angle. The medium and large displays measured 19.1 by 14.0 cm (20° by 15°) and 34.3 by 25.4 cm (36° by 27°), respectively. With all changes to display size, the text and icons located within the display also changed proportionately.

Procedure

Participants were asked to complete two tasks, namely flight control and hazard surveillance. In the flight control task, pilots were asked to maintain a target flight level of 15,000 feet and a north-up heading. Vertical and lateral maneuvers were made with a two-axis joystick. While altitude information could be determined from the digital readout in ownship's data tag and from the attitude directional indicator, heading information could only be deduced from the orientation of the aircraft icon representing ownship. Participants were also asked to maintain a separation of 5,000 feet from a lead aircraft by increasing and decreasing their airspeed. The target separation distance of 5,000 feet was depicted in a scale that was located on the bottom right-hand corner of the display.

While performing the flight control task, pilots were also asked to monitor the airspace for changes in the heading, altitude, or airspeed of traffic aircraft and weather systems. These hazard changes occurred randomly every 15 to 75 seconds. Pilots were asked to identify changes with a key press and verbal description of the change (e.g., "Aircraft C changed heading"). While altitude and airspeed events could only be detected by noting the changes in the hazard's digital data tag, heading changes could be detected by viewing the heading information located within the data tag or by noting a change in the movement of the hazard. Participants completed one practice trial and six experimental trials. Each trial lasted six minutes and the experimental session lasted for about one hour.

Experimental Design

Display size was counterbalanced and manipulated as a within-subjects variable. For the task of flight control, measures included lateral and vertical root mean squared (RMS) error and mean absolute error in tracking the target separation distance from the lead aircraft. Change detection performance was evaluated with measures of response time and accuracy. Surveillance performance was also assessed through measures of percent dwell time and mean dwell duration to three designated display regions.

Results

Change Detection Performance

On average, pilots detected 12.2% of changes with a latency of 18.0 s. Change detection accuracy and response time were both evaluated in a one-way repeated measures ANOVA as a function of display size. These analyses revealed no significant effect of display size on either accuracy ($p > 0.10$, $\phi = 0.48$) or response time ($p > 0.10$, $\phi = 0.26$). Independent of display size, the influence of change eccentricity on detection performance was also assessed by evaluating by accuracy and response time as a function of the distance of the event from ownship, which was assumed to be the focus of attention. This analysis yielded a significant correlation between change eccentricity and detection accuracy ($r = -0.49$, $p < 0.01$). Thus, detection accuracy was significantly reduced as the change occurred at an increasingly greater distance from ownship, independent of the relevance of the event to ownship's safety.

Given that display size had no effect on surveillance performance, the analyses suggest that performance was degraded as changes occurred further from the center of the display. However, display enlargements, which served to further increase the distance between the center of the display and the display perimeter, did not amplify this effect. This latter finding suggests that pilots were *strategically compensating* for display enlargement by widening their scanning patterns. This can be confirmed by examining the eye movement data.

Eye Movement Data

Eye movement data was collected and assessed as a function of percent dwell time and mean dwell duration in three designated display regions, as shown in Figure 2. The *ownship* display region included ownship, a lead aircraft, and the attitude directional indicator. The *midrange* display region included the area immediately surrounding the ownship region. The most peripherally located region was the *outer* display area, and included the area of the map around the display perimeter.

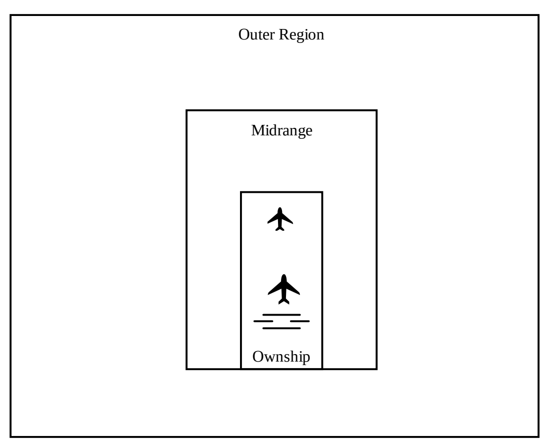


Figure 2. *Display regions or areas of interest.*

Using a median split, pilots were also grouped into high and low change detection performers. While head movement data was collected, participants rarely used head movements to access information located on the display. Consequently, these data will not be discussed.

Percent dwell time and mean dwell duration were assessed in Display Size X Display Region X Change Detection Performance mixed ANOVAs. Percent dwell time analyses revealed a significant main effect of display region, with participants allocating the greatest proportion of attention to the ownship region ($F(2, 20) = 56.13$, $p < 0.001$). Interestingly, the outer region received a significantly greater proportion of attention than the midrange region, and this effect increased with display enlargement from the small to medium display ($F(4, 40) = 4.12$, $p = 0.007$). There was also a shift in attention away from the midrange region to the ownship region from the small to medium display, suggesting that pilots needed to foveate the ownship area to gain flight control information when the display was enlarged. Interestingly, there was no significant effect on percent dwell time for display enlargement from the medium to large display, suggesting that pilots strategically compensated for display size in their scanning patterns. These relationships are depicted in Figure 3.

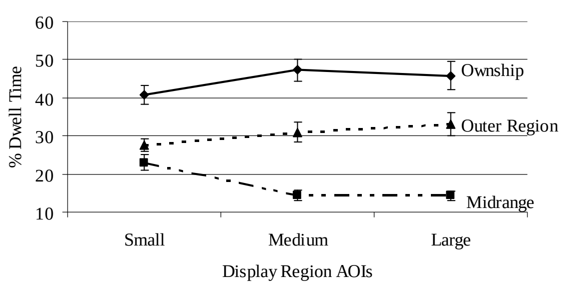


Figure 3. *Significant interaction of display region and display size.*

While a significant proportion of attention was allocated to the outer display region, this region was also the largest in area. Thus, when percent dwell time was normalized by a measure of percent/cm^2, this measure declined monotonically from the ownship region of the display to the midrange and outer regions ($F(2, 20) = 274.0$, $p < 0.001$). Thus, while the outer region received more total attention than the midrange region, the allocation was more sparsely distributed across display area. These findings support the performance analyses that revealed a decrease in change detection accuracy for events in the more eccentric outer display region.

Mean dwell duration in each of the three regions was also examined to determine if the differences found in percent dwell time were due to more scans or longer fixations within each display region. The mean dwell duration data, examined as a function of display size and display region are plotted in Figure 4. The analyses revealed a significant main effect of display region ($F(2, 20) = 132.8$, $p < 0.001$), with dwells in the ownship region lasting more than three times the length of those in the midrange and outer display regions. This finding reflects the need to access information about flight control from this region.

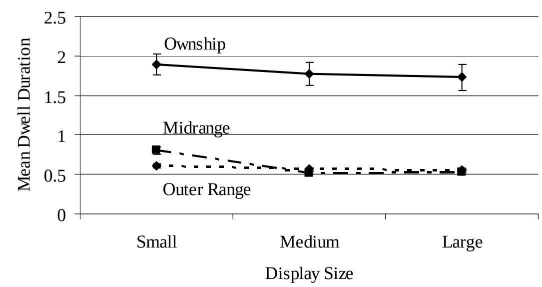


Figure 4. *Mean dwell duration as a function of display size in each of the three display regions.*

Analyses also indicated that dwell duration for the midrange and outer display regions did not significantly differ, at least for displays that were medium or large in size. Thus, the difference in percent dwell time for the midrange and outer display regions was *not* due to a difference in dwell duration, but rather can be attributed to a greater number of visits. These findings provide additional support for the *strategic compensation* model of surveillance, suggesting that pilots fixated the outer region more frequently and with longer scans than the midrange area.

Finally, surveillance was assessed as a function of change detection performance. These analyses indicated that good performers allocated a greater proportion of attention to the outer display region, while attention for the low performers was more solely concentrated to the ownship region ($F(2, 20) = 5.84$, $p = 0.01$). This difference strengthened when displays were enlarged from small to medium ($F(4, 40) = 2.14$, $p = 0.09$), though the interaction was only marginally significant. High performers were also found to have shorter dwells than low performers, though only for the ownship region ($F(2, 20) = 3.73$, $p = 0.04$). Thus, these data suggest that high performers were particularly skillful at allocating attention away from the ownship region to the more peripheral regions of the display. This *strategic compensation* was particularly apparent with the medium and large displays.

Discussion

The present study was designed to examine two strategies of surveillance in response to display enlargement. The first model, *effort conservation*, posited that pilots would be unable to sustain extended surveillance patterns, opting instead to conserve scanning effort by concentrating on the central portions of the display (Enoch, 1959). Some evidence for the effort conservation approach was found in the eye movement behavior of the poor change detection performers, who spent too long focusing on the proximal tracking task and failed to allocate attention to the outer display regions to detect distant events, particularly for large displays. Despite this evidence, change detection performance for the group as a whole was unaffected by display enlargement, suggesting instead that pilots adopted the strategic compensation model.

The *strategic compensation* model posited that pilots would adapt to enlargements in display size by widening their scanning patterns to monitor even the most peripherally located display regions. Evidence for the strategic compensation model was found in the scanning data for both high performers and that of the overall participant group. The overall analyses indicated that, while the outer region received the smallest proportion of attention per square centimeter of display area, this proportion did not decrease with display enlargement. In fact, for the high performers, this proportion *increased* when the display size was enlarged from small to medium. Thus, pilots were able to widen their scanning patterns without a performance cost (Teichner & Mocharnuk, 1979).

A final form of strategic compensation was evidenced in the elevated values of percent dwell time and mean dwell duration for the midrange region in the small display. It is our belief that, when the display was presented in the small format, pilots were able to fixate in the middle display region while maintaining the ownship region within the useful field of view. Thus, with the small display pilots might have chosen a strategy to fixate more often in the middle region, knowing that by doing so, they did not need to temporarily abandon the flight control task.

While the strategic compensation strategy used by pilots sustained change detection performance across display sizes, it likely did not come without cost. Any widening of the scanning pattern with an enlargement in display size would also produce an increased demand for resources (Recarte & Nunes, 2002). To the extent that the scanning task becomes more difficult because the display becomes excessively or concurrent tasks are added, the pilot may turn to an effort conservation approach to cope with the increased demands. This may primarily occur when display size is increased to such a large degree as to induce head movements, though this threshold was not examined in the present study. Additionally, pilots represent a population who has been thoroughly schooled on the importance of scanning displays and instruments, despite the extra effort that must be employed to do so. Consequently, care should be taken in extending these data to other domains whose operators do not share this characteristic.

Conclusions

Despite the increase in effort associated with monitoring large displays, pilots demonstrate adaptiveness by widening and enlarging scan patterns in order to access information needed for safe flight. At a practical level, the results suggest that displays of this sort can be enlarged up to thirty degrees of visual angle without much *performance* cost, though workload will be increased. Care should also be taken to ensure such an enlargement will not simultaneously hinder additional tasks supported by the display.

Acknowledgments

This work was supported by a grant from the NASA Ames Research Center (NASA NAG 2-1535), for which Dr. David C. Foyle served as the technical monitor. The authors also wish to acknowledge Sharon Yeakel and Roger Marsh for their programming efforts.

References

Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology, 14,* 468-469.

Carpenter, S. (2001). Sights unseen. *APA Monitor, 32,* 54-57.

Enoch, J. M. (1959). Effects of the size of a complex display upon visual search. *Journal of the Optical Society of America, 49,* 280-286.

Kroft, P., & Wickens, C. D. (2003). Displaying multi-domain graphical database information: An evaluation of scanning, clutter, display size, and user interactivity. *Information Design Journal, 11,* 44-52.

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review, 4,* 501-506.

Muthard, E. K., & Wickens, C. D. (2002). Change detection after preliminary flight decisions: Linking planning errors to biases in plan monitoring. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting.* Santa Monica, CA: Human Factors and Ergonomics Society.

Pringle, H. L., Irwin, D. E., Kramer, A. F., & Atchley, P. (2001). The role of attentional breadth in perceptual change detection. *Psychonomic Bulletin & Review, 8,* 89-95.

Recarte, M. A., & Nunes, L. (2002). Mental load and loss of control over speed in real driving: Towards a theory of attentional speed control. *Transportation Research Part F: Traffic Psychology & Behavior, 5,* 111-122.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8,* 368-373.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53,* 245-277.

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition, 7,* 1-15.

Teichner, W. H., & Mocharnuk, J. B. (1979). Visual search for complex targets. *Human Factors, 21,* 259-275.

Wickens, C. D. (1992). *Computational Models of Human Performance* (Technical Report No. ARL-92-4/NASA-A3I-92-1). Savoy, IL: University of Illinois, Aviation Research Laboratory.

OPERATORS' TIME PERCEPTION UNDER STRESS

R.V. Nayeem[1], **T. Oron-Gilad**[1&2], & **P. A. Hancock**[1&2]
Department of Psychology[1] and [2]Institute for Simulation and Training
University of Central Florida

Time perception is extremely important to the understanding, design and use of complex aviation systems. This experiment focused on differences in time estimation, flight performance, and monitoring tasks. In a between-subjects experiment, participants navigated through a flight scenario while monitoring a switch and listening to white noise at either 55dBA or 85dBA. Flight performance data and monitoring data were collected throughout the task. Participants also completed the NASA-TLX and the DSSQ-S. Statistical analyses showed that the noise condition did not significantly affect workload, monitoring abilities, task completion and subjective stress questionnaires for the dual task. However, the 85dBA condition significantly affected prospective time estimation. These results suggest that the dual task was not demanding enough, and the stress was not adequate to push participants out of the comfort range and experience a performance decrement.

Introduction

The temporal domain is stress-sensitive in a similar manner to the spatial domain and comparable narrowing occurs, resulting in distortions of perceived time (Hancock & Weaver, 2005). Understanding time perception and time distortion under stress is of high importance for operating complex systems. Time perception affects how operators react to visual, auditory and tactile alarms. If time estimation is inaccurate due to stress, the alarm may go unattended for a critical length of time which exceeds the time available for solution. Time estimation is affected by the stress and mental effort that operators experience. If stress conditions are sufficiently high, such conditions will induce time distortion (Block, Zakay, & Hancock, 1999). This has been seen in Eastern Airlines Flight 401 in the Everglades where the crew got fixated on a landing light, and in the infamous John Denver fatal aircraft accident.

One possible solution to this problem is adaptive automation. Adaptive function allocation or dynamic automation is an approach in which control of tasks dynamically shifts between humans and machines, and is an alternative to traditional static allocation in which task control is assigned during system design and remains unchanged during operations. With adaptive automation, if operators are under too much stress to respond optimally to the aircraft, tasks will automatically get shifted to the automation rather than the operator. If automation states are switched too quickly, operators can become confused and decision making can be degraded, and allocation induced oscillations car occur (Hancock & Scallen, 1996). Also, when a trigger is operator-driven, the added workload and decision making can hinder performance (Hildebrandt & Harrison, 2002). Using dynamic function scheduling, an operator can prioritize re-allocation by assigning tasks both a temporal value, when the task should be re-allocated along the system timeline, and a qualitative value, what contribution the task makes to the system goal and the quality of the eventual solution. The Hancock and Warm (1989) model shows different zones in which a person can be when stressed.

The purpose of this study was to investigate differences on time estimation due to differences in the audible level of white noise, while performing a flight navigation task and a dual monitoring task. Stress was manipulated by task complexity, time constraints, and the addition of one of two levels of white noise: standard high frequency white noise at either 55dBA or 85dBA. This study attempted to investigate time distortion associated with stress and monitoring tasks in a basic flight navigation task. It was anticipated that participants would be pushed out of the comfort zone (Hancock & Warm, 1989) and would experience performance decrements to workload, monitoring abilities, subjective stress questionnaires, and prospective time estimation. However, given the relatively simple nature of the task and the qualifications for participation, it was expected that the noise would not affect task completion.

Method

Participants. 32 pilots, 28 male and 4 female students enrolled in psychology and human factors courses at Embry-Riddle Aeronautical University. The mean age was 21.19 (SD=3.61). Participants were required to have at least a private pilot's license. Total hours as PIC

[1] Corresponding Address: Razia V. Nayeem, Department of Psychology, PO Box 161390, University of Central Florida, Orlando, FL 32816-1390 E-mail: rnayeem@ist.ucf.edu

ranged from 40 to 700, with a mean of 133.1 (SD=134.1).Students who chose to participate in this research effort were compensated via extra credit. The participants all filled out personality questions as well. On a rating scale of 1-5, they uniformly rated themselves as able to pace themselves to get things done (M=4.0, SD=.6), able to handle stress (M=2.3, SD=.9), as moderately methodical people (M=2.4, SD=.8), and as having fast-paced lives (M=3.3, SD=1). About half of the participants responded that they waste a lot of time before settling down to work, and about half responded that they do not (M=3.9, SD=.7).

Apparatus and Tasks. The experimental setup consisted of two independent tasks generated in the AirBook® by Simigon F-16 flight simulator: a) the monitoring task and b) navigation task. The flight simulator was installed on a Dell laptop model # PP08L, with a 17" screen. Static white noise was emitted through headphones plugged into the computer speakers.

White Noise. There were two levels of white noise – 55dB and 85dB. The sound was delivered via Maxell headphones, model # HP/NC-II, and was measured before each participant via a Sper Scientific sound meter, model # 840029.

Flight Controls. Participants manipulated the flight environment using a GF yoke, model # G60503A.

Navigation Task. The navigation task objective was to take off from an airport, navigate to three waypoints, and land at the initial airport, as shown in Figure 1. A skyscraper marked each waypoint. After the participant crossed over the waypoint at 5000 feet MSL, the avionics of the aircraft changed to guide the participant to the next waypoint. When a participant crossed all three waypoints, the task was considered successfully complete. The waypoints were strategically placed throughout Arizona terrain. Two different navigational missions were used for the practice trial and the dual task trial so that participants would not become too familiar with any one mission. The navigation task difficulty did not vary from mission to mission. Additionally, at the end of the second mission, participants were asked to estimate how long they spent completing the dual task.

Monitoring Task. The monitoring task consisted of periodically checking the master switch in the aircraft console. The switch was set to turn to the "off" position 30 seconds after takeoff, and 30 seconds after the participant crossed the second waypoint. Switch states are shown in Figure 2. Reaction time and accuracy were recorded for the monitoring task. If the participant did not turn set the switch to the "off" position after the first change, the switch remained in this position for the entire flight. Similarly, if a participant did not set the switch to the "off" position after the second change, the switch remained in this position for the remainder of the flight. If participants did not cross the second waypoint at all, or higher than 5000 feet, the switch remained in the position it was prior to these events.

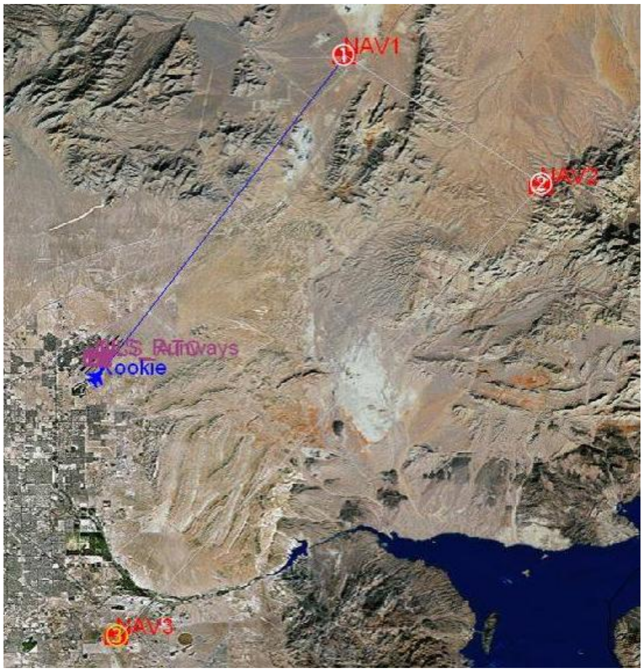


Figure 1. *The navigation route in the AirBook® environment. The objective of the navigation task was to take off from an airport, navigate to three waypoints, and land at the initial airport.*

(a) (b)

Figure 2. *The Monitoring task required pilots to respond to the status of the master switch. In (a) the master switch is armed in (b) it is off.*

Time Estimation. During the flight participants were required to estimate the duration of the task they performed (Zakay & Block, 2004). This prospective duration estimation can serve as a secondary task for workload measurement (Block, Zakay, & Hancock, 1999).

Subjective Measures. Questionnaires included a demographic questionnaire, the DSSQ-S (Matthews, et al., 1999), and a computer version of the NASA-TLX (Hart & Staveland, 1988).

Procedure. After a brief description of the experiment by the experimenter, participants signed an informed consent form and completed a demographic questionnaire. Once informed consent was received, participants were familiarized with the AirBook® navigation task. They were given written and verbal instructions on how to maneuver through the AirBook® environment using the keyboard and the yoke provided for flight inputs. Then, participants had a five-minute practice. After the practice session, participants completed an initial NASA-TLX to be used as a baseline, and the DSSQ-S pre. Participants were then given the option to take a five-minute break. After this, participants were given brief instructions, and the dual task began. Participants navigated through a predefined route while performing the monitoring task for ten minutes. Performance on the monitoring task was recorded. In addition, during the dual task, participants received white noise through headphones. Half of the participants received noise at 55 dB, and the other half received noise at 85 dB. After completion of the dual task, participants estimated how long they were doing the dual task, completed a second NASA-TLX and the DSSQ-S post and were completely debriefed, having all their questions answered.

Results

All analyses were conducted using SPSS 11.5 for Windows. All alpha levels were set to .05.

Flight Performance

Navigation Task. Twenty-three of the 32 participants (72%) successfully completed the navigation task. Seven of those who did not complete the task had lower total flight hours than the mean. The other two had 140 and 700 flight hours, respectively. A one-way ANOVA on task completion with noise as a between subject variable was not significant, $F(1,30) = 0.146$, $p = 0.705$.

Monitoring Task. The mean reaction time for alarm 1 was 100.28 seconds (SD=186.36), and the mean reaction time for alarm 2 was 240.96 (SD=93.07). Twenty-eight of the 32 participants responded to alarm 1, and 23 of the 32 participants responded to alarm 2. All participants that failed to respond to alarm 1 had less flight hours than the mean except one participant, who had 700 hours. All participants that failed to response to alarm 2 had less flight hours than the mean except two participants, who had 140 hours and 700 hours.

A paired-samples T-test was conducted on the reaction times for alarm 1 and alarm 2 revealed that there was a significant difference in response time, $t(22) = -11.6$, $p < .0005$, Cohen's $d = -3.3$, effect size $r = -.86$. This difference is probably due to the fact that as the flight progressed participants paid less attention to the monitoring task. Two one-way ANOVAs on response to the first and the second alarm with noise as a between subject variable were not significant, $F(1,30) = 1.111$, $p = 0.3$, $F(1,30) = 0.146$, $p = 0.705$, respectively. The mean time to turn off alarm 1 in the 55dBA condition was 71.69 seconds (SD=138.9), and 128.88 seconds (SD=225.3) in the 85dBA condition. The mean time to turn off alarm 2 in the 55dBA condition was 208.75 (SD=94.5) seconds, and 276.09 (SD=81.4) seconds in the 85dBA condition. Two one-way ANOVAs on response time to alarm 1 and alarm 2 with noise as the between-subject variable were also not significant, $F(1,30) = 0.747$, $p = 0.394$, $F(1,21) = 3.322$, $p = 0.083$, Cohen's d = , effect size r = , respectively.

Stress and Workload.

NASA-TLX. A repeated-measures ANOVA before the task (M=52.40, SD=15.68) and after the task (M=53.25, SD=20.58) with noise as a between subjects variable revealed that the overall pre- and post- TLX measures were not affected by noise, $F(1,30) = 0.113$, $p = .739$.

A one-way ANOVA of the TLX scores after completion of the task indicated that there were no significant difference in NASA-TLX scores between the 55dBA condition and the 85dBA condition, $F(1,30) = 0.908$, $p = .348$. Additionally, the six sub-scales were analyzed for significance and effect size using one-way ANOVAs. The statistical results can be seen in Table 1.

NASA-TLX after completion of task	55dBA	85dBA	F	df	p	Cohen's d	Effect Size r
Global TLX	M=55.7 SD=21.5	M=50.8 SD=20.0	0.908	(30)	.348	.24	.12
Mental Demand	M=66.6 SD=24.4	M=58.4 SD=24.3	1.256	(1,30)	.271	.33	.16
Physical Demand	M=30.0 SD=18.3	M=35.5 SD=24.3	1.533	(1,30)	.225	-.25	-.12
Temporal Demand	M=56.9 SD=21.1	M=41.6 SD=25.5	0.058	(1,30)	.811	.65	.31
Performance	M=45.0 SD=33.2	M=36.3 SD=18.5	2.837	(1,30)	.103	.31	.15
Effort	M=70.3 SD=22.7	M=60.3 SD=30.0	2.034	(1,30)	.164	.38	.18
Frustration	M=35.6 SD=26.6	M=39.4 SD=30.9	0.244	(1,30)	.625	-.13	-.06

Table 1. *Analysis of the differences in TLX scores after completion of the dual task by noise condition .*

Time Estimation. The actual duration of the task was 10 minutes. The estimated times had a mean of 7.7 minutes (SD=3.0). Zakay (1998) suggested that underestimation is more likely to occur when time estimation is not the primary task at hand. As more attention is being focused on temporal information processing, more time signals are processed and the judgment is more likely to be accurate. Allocating fewer attentional resources to temporal information processing causes fewer time signals to accumulate, resulting in a decrease in the estimated duration (Zakay & Block, 2004).

Time estimates were translated into the duration judgment ratio (DJR; Block, Zakay, & Hancock, 1999). Correlations were calculated for the DJR (M=77.3, SD=29.7) and the NASA-TLX score for participants in the 55dBA condition after the dual task, and the DJR and the NASA-TLX score for participants in the 85dBA condition to determine whether or not time estimates were correlated with noise condition. The correlation between the DRJ and the 55dBA condition was not significant, r = .164, p = .543. However, the correlation between the DJR and the 85dBA condition was significant, r = -.57, p<.05. This suggests that the 85dBA noise condition affected the DJR.

DSSQ-S. The DSSQ-S is scored into engagement, pre (M=21.3, SD=3.7) and post (M=22.7, SD=4.0), distress, pre (M=8.4, SD=5.8) and post (M=7.0, SD=4.4), and worry, pre (M=6.6, SD=4.6) and post (M=5.8, SD=4.3). Paired-samples t-tests were performed for engagement pre-post, distress pre-post, and worry pre-post, t(31) = -1.97, p = .058, Cohen's d = -.33, effect size r = -.17, t(31) = 1.48, p = .15, Cohen's d = .27, effect size r = .13, and t(31) = 1.45, p = .157, Cohen's d = .18, effect size r = .09, respectively. The difference in the pre and post engagement scores was not statistically significant, though it was very close. However, the effect size was not large enough to warrant more examination.

Repeated-Measures. ANOVAs were performed for each sub-scale with noise as the between subjects variable, F(1,30) = 3.79, p = .061 Cohen's d = .37, effect size r = .18, F(1,30) = 2.15, p = .153, Cohen's d = -.23, effect size r = -.11, and F(1,30) = 2.104, p = .157, Cohen's d = .06, effect size r = .03, respectively. Thus, noise was not a moderating condition for the DSSQ-S sub-scales.

Correlations were performed for the DSSQ-S measures before and after the task with noise condition and pre- and post- TLX global and performance scores using the Spearman ranking coefficient. For the correlations computed, the only ones that were statistically significant at the .05 level was the correlation between the post-task NASA-TLX global scores and the DSSQ-S post-task distress measure, ρ = .384, p = .03, pre-task distress and pre-task mental demand, ρ = .439, p = .012, pre-task distress and pre-task frustration, ρ = .434, p = .013, and post-task distress and post-task effort, ρ = .377, p = .033. These correlations indicate that there is a strong relationship between overall workload, post-task effort and pre-task performance with distress after the task. Additionally, these correlations indicate that there is a strong relationship between pre-task mental demand and pre-task frustration with pre-task distress.

Discussion

The majority of the pilots (71.9%) were able to complete the navigation task without any difficulties. However, since the workload estimates indicated that participants were not experiencing a heavy workload, one would think that close to 100% of the participants would have finished the task. Of the participants who could not complete the task, 4 were in the 55dBA condition and 5 were in the 85dBA condition. This suggests that task completion was not strongly affected by noise condition. Thus, task completion must be mediated by other moderators. When looking at the DSSQ-S engagement scores after the task, the mean was 22.6 compared to a maximum possible score of 28. Thus, the average score was only approximately 80% of the possible score. This suggests that participants were simply not engaged enough in the task to complete the task. To remedy this in the future, the navigation task will be comprised of more tasks that are shorter and a more demanding navigation scenario.

Performance on the monitoring task varied among participants, but this variation did not seem to correlate with the noise condition. The data suggests that some people simply ignored the alarms, possibly because there were no consequences, and the changes of the switch were very subtle. To better this task in the future, participants will be given a certain amount of time to comply with the alarm, and after that time expires, a visual alert will be shown to direct attention to the switch.

Interestingly, the mean time to turn off Alarm 2 in both conditions was more than twice the time it took to turn off Alarm 1. This suggests that participants were not monitoring the switch as closely throughout the flight as they were in the initial phases of flight. This is consistent with other monitoring task experiments, in which participants experience a performance decrement after a certain period of time.

The actual results were somewhat different from the expected results, in that workload, monitoring abilities, and subjective stress questionnaires were not affected by noise. However, as expected, task completion and time estimates were affected by noise. In the Hancock and Warm (1989) model, this means the stress was not sufficient to push participants out of the comfort zone, which is reflected in the lack of performance decrement in the flight parameters. Future work will address the issues that were problematic in this study.

Acknowledments

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621, Dr. Elmar Schmeisser, Technical Monitor, P.A. Hancock, Principal Investigator. The views expressed in this work are those of the authors and do not necessarily reflect official Army or Department of Defense policy.

References

Block, R.A., Zakay, D. & Hancock, P.A. (1999). Developmental changes in human duration judgments: a meta-analytic review. Developmental Review, 19(1), 183-211.

Chignell, M.H., & Hancock, P.A. (1987). A comparison of mental workload and available capacity in complex person-machine systems: A fuzzy-set theoretic approach. In: W. Karwowski and A. Mital (Eds.). Applications of Fuzzy-Set Theory in Human Factors. (pp. 271-288), Amsterdam: Elsevier Publishers.

Gluckman, J. P., Morrison, J. G., & Deaton, J. E. (1991). Complex task performance as a basis for developing cognitive engineering guidelines in adaptive automation. Proceedings of the Human Factors Society 35th Annual Meeting, Santa Monica, CA: Human Factors Society

Hancock, P.A. & Warm, J.S. (1989). A dynamic model of stress and sustained attention. Human Factors, 31, 519-537.

Hancock, P. A., & Weaver, J. L. (2005, in press). Temporal distortions under extreme stress. Theoretical Issues in Ergonomic Science.

Hildebrandt, M. & Harrison, M. (2002). The temporal dimension of dynamic function allocation. Paper presented at the Eleventh European Conference on Cognitive Ergonomics.

Parasuraman, R. & Mouloua, M. (1996). (Ed.). Automation and human performance. Lawrence Erlbaum: Mahwah, N.J.

Scallen, S.F. & Hancock, P.A. (2001). Implementing adaptive function allocation. The International Journal of Aviation Psychology, 11(2), 197-221.

BEYOND AUTOMATION SURPRISES: A SIMULATOR STUDY OF DISTURBANCE MANAGEMENT ON HIGHLY AUTOMATED FLIGHT DECKS

Mark I. Nikolic
The Ohio State University
Department of Industrial and Systems Engineering
Columbus, OH

Nadine B. Sarter
The University of Michigan
Center for Ergonomics
Ann Arbor, MI

Error prevention alone will never be sufficient for improving safety in complex high-risk systems, such as aviation. This approach needs to be combined with better support for error and disturbance management which, in turn, requires an improved understanding of current strategies for coping with errors and the resulting disturbances to the flight. The present research has sought systematic empirical evidence to expand our understanding of the disturbance management process on modern flight decks. A simulator study was conducted with twelve B747-400 airline pilots in order to examine (the effectiveness of) their strategies for diagnosing and recovering from disturbances, and the impact of current automation design on these processes. Pilots flew a one-hour scenario (with a confederate copilot) which contained challenging events that probed pilots' knowledge of, and proficiency in, using the autoflight system. A process tracing methodology was used to analyze and identify patterns in strategies across pilots. Overall, pilots completed the scenario successfully but varied considerably in how they coped with disturbances to their flight path. Our results show that aspects of feedback design delayed the detection, and thus escalated the severity, of a disturbance. Diagnostic episodes were rare due to pilots' knowledge gaps as well as time-criticality. Our findings can inform the development of design and training solutions to observed difficulties with error and disturbance management in a variety of domains.

Introduction

Human error is cited as the cause, or a contributing factor, in the majority of aviation incidents and accidents (e.g., Boeing, 1994). Yet the very low accident rate in this domain illustrates that aviation exhibits a strong degree of error resilience thanks to pilots' successful management of their errors and associated disturbances to the flight. In highly complex, dynamic, and event-driven domains, such as aviation, operators often need to manage consequences of breakdowns in human-machine performance that interact, cascade and escalate over time while continuing to maintain the ongoing process (such as flying the airplane). This activity can be characterized as disturbance management, since, from a practitioner's perspective, the potentially negative system effects of an error are more relevant than the error per se.

Disturbance management refers to the activity of diagnosing the underlying source(s) of a disturbance (i.e., a deviation from a desired state) in parallel with coping with the disturbance itself by maintaining the integrity and goals (i.e., efficiency, safety) of an underlying dynamic process (Woods, 1988). In the aviation domain, for example, a pilot needs to diagnose the source (for example, an erroneous input to the FMS) of an observed disturbance (such as a deviation from the flight path) and cope with the disturbance (by bringing the airplane back on course) while maintaining the integrity of the underlying process (i.e., while continuing to fly the airplane). While disturbance management is usually discussed in the context of system faults, the same activities tend to be involved in handling the consequences of breakdowns in the interaction between humans, machines, and the complex dynamic environment in which they collaborate. We will therefore use the term "disturbance management" to refer to pilots' efforts to cope with the effects of automation-related erroneous actions and assessments.

Despite the importance of disturbance management for system safety, few studies have examined its components in real-world dynamic environments (for some examples, see Klinect et al., 1999; Woods, 1984). The majority of work in this area has focused on error detection, leaving unanswered questions about the other stages of disturbance management (i.e., diagnosis and recovery). Earlier studies suggest that diagnosis does not necessarily occur or precede recovery during dynamic disturbance management (Kanse and Schaaf, 2001). Furthermore, an examination of how technological tools shape disturbance management seems to be missing from most earlier efforts.

.In the context of pilot-automation interaction and performance breakdowns on modern flight decks, our goal was to determine whether, and under what circumstances, pilots attempt to diagnose before they respond, and to what extent diagnosis is required for successful disturbance handling. Another objective was to examine the range of recovery strategies used, especially when they are influenced by the design of flight deck automation.

Methods

As the final step in a research program that included jump-seat observations, a flight instructor survey, and an incident database analysis, a high-fidelity simulator study was conducted with type-rated airline pilots in order to examine error and disturbance management in a semi-controlled full-mission flight simulation context.

Participants

Pilot volunteers were recruited from two major U.S. carriers and one airplane manufacturer. Twelve type-rated Boeing 747-400 pilots (11 current, 1 recently retired; mean hours on type = 3837.75, SD = 2478) participated in the study and were paid $100 for their involvement.

Simulator

The simulation was conducted on a fixed-base 747-400 flight simulator. The 747-400 is a highly automated four-engine long-haul passenger aircraft. The simulator was equipped with fully functional displays and control interfaces. An Evans & Sutherland ESIG 3350 image generation system rendered a panoramic out-of-window visual scene which covered 45° horizontally and 34° vertically for each pilot.

Procedure

After briefing the flight with the experimenter and reviewing all flight-related paperwork, the participating pilot joined the confederate pilot in the simulator. The confederate knew the purpose of the study, occupied the right (co-pilot) seat, and helped ensure that scenario events occurred as designed. The confederate pilot was instructed not to be overly proactive in helping participating pilots detect their errors. However, he was instructed to intervene (by directing the participant's attention) if the detection delay jeopardized the experimenter's likelihood to observe a recovery. The confederate was also asked to elicit pilots' reasoning about problems by asking relevant questions to expose the pilot's intentions and reasoning. Interactive air traffic control was provided by the experimenter/observer to help ensure the proper evolution of the scenario by issuing planned and improvised clearances. After reviewing the planned route and the current state of the aircraft, the scenario began in-flight with the aircraft level at 9000 feet, during the initial climb-out phase. The scenario ended once the aircraft landed at Los Angeles and came to a complete stop on the runway. The pilot then remained in the simulator cab and was debriefed by the experimenter for another 30-60 minutes.

Scenario

All participants flew the same one-hour daytime scenario from San Francisco to Los Angeles in the role of pilot-in-command. Weather throughout the scenario was clear with minimal winds. Based on data gathered from our earlier survey, observations, and consultations with domain experts, several scenario events were designed that created a high probability of observing automation-related disturbances by placing heavy knowledge and attentional demands on pilots resulting in the potential for breakdowns in human-machine communication and coordination. Since errors and disturbances were not introduced through experimenter-induced system failures or unrealistic clearances, they were not necessarily observed for each pilot on each event.

Selected Scenario Events

Because of space limitations, this paper will present results from two of the events that were used in the scenario.

LNAV Capture. After crossing PESCA, ATC instructed the aircraft to continue on a 140 degree heading instead of turning left to continue on the flight plan. As a result, the aircraft will not physically cross the next two waypoints that are programmed into the Flight Management Computer (FMC) and are kept in the route. Thus, if the route is not reprogrammed by the pilot, the autopilot will attempt to return to these waypoints and result in unwanted aircraft behavior when the pilot attempts to rejoin the course by activating the LNAV mode.

VNAV ALT Mode. In order to begin an automated descent, the autoflight system must be in the 'VNAV PTH' mode. However, in our scenario, the automation was likely to enter the 'VNAV ALT' mode due to cruise altitude changes given by air traffic control. If the pilot does not actively change the mode back to 'VNAV PTH' (typically, by changing the cruise altitude in the CDU interface of the Flight Management System and then pushing the altitude knob), the aircraft will not descend as expected at the top-of-descent (TOD) point, and may potentially miss an altitude target. This event could elicit a mode error due to either incomplete system knowledge or a monitoring breakdown, and could have resulted in an altitude violation if it was not detected and corrected in a timely manner.

Data Collection and Analysis

Multi-angle video and audio recordings were made to assist in recreating verbal and behavioral protocols. This information was supplemented by an observer, who sat directly behind the pilots in the simulator cab and noted pilot responses to events. Upon completion of the scenario, the participating pilot was debriefed by the experimenter in order to review and clarify any ambiguities about his scenario performance and to probe participants' knowledge of the automated flight system. These sources of data were combined to form a coherent process trace (Woods, 1993) of participant behavior which can be compared across participants as well as to canonical or "standard" recovery paths for each event.

Results

All twelve pilots completed the scenario for this study "successfully" in the sense that they all made a safe landing. However, every pilot struggled at some point with handling events during the simulated flight, and every scenario unfolded in a unique way because pilots used a variety of strategies for managing events and recovering from disturbances. When possible, a canonical solution path was defined by a subject matter expert for the event. This path represented the most efficient but not necessarily the only correct or successful sequence of pilot actions for the event. It provides a single frame of reference from which to compare performance across pilots.

LNAV Capture

This event examined how pilots recovered their original course after an air traffic control clearance caused them to bypass two of the waypoints on the original route. As a consequence, the FMS continued to consider them as "active" (i.e., as valid targets) since the airplane never came close enough for them to be removed by the automation's logic. As a result, pilots who re-activated LNAV to resume the course without first modifying the route in the CDU caused the airplane to turn off-course to a 090 heading instead of the 070 heading as instructed by ATC.

All pilots eventually managed to recover their course, although minor deviations occurred for two pilots (see Figure 1). After receiving the ATC clearance to intercept their normal course via a 070 heading, all pilots made the initial turn using the HDG SEL mode (a lateral mode at a low level of automation). The recovery processes from that point on fall into two categories. One group of five pilots reprogrammed the route prior to engaging the LNAV mode (represented on the lower half of Figure 1). A second group of seven pilots activated the LNAV mode without updating the original route in the FMS. In general, the group that reprogrammed the route first was more aware of the current state and logic of the automation. These pilots recognized that the FMS route contained "stale" information that was no longer applicable to the new context.

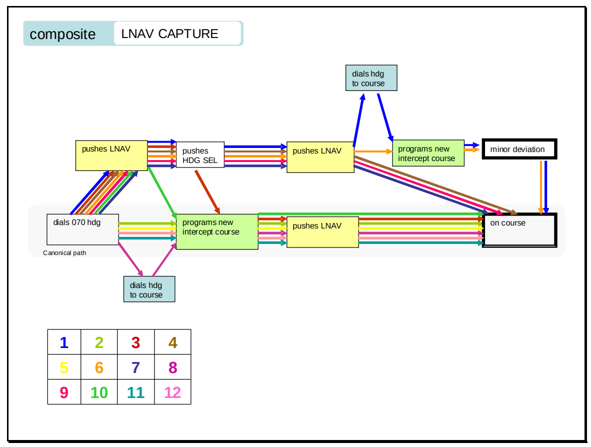


Figure 1. *Composite of abstracted solution paths for LNAV event. Each line represents one pilot and is color-coded for all pilots.*

For the second group of pilots, it was not immediately obvious that there was a problem because the incorrect FMS route produced aircraft behavior that was initially consistent with pilot expectations. Since the FMS believed that the floating waypoint (a turn to a 090 degree heading) was the current target, activation of the LNAV mode resulted in a turn in the expected direction (left) but not to the assigned heading of 070 degrees. This confirming cue initially masked the problem and led some pilots to assume the aircraft was on the correct course.

Six of the seven pilots in the second group (top portion of Figure 1) recovered the correct heading by reverting back to the HDG SEL mode, after detecting either the unexpectedly rapid engagement (or "capture") of the LNAV mode or the subsequently incorrect heading of 090 degrees that was commanded by the autopilot. Of the six pilots that recovered with HDG SEL, one of them detected the active waypoint mismatch at this point and reprogrammed the route, while five of them reattempted to engage the LNAV mode, again, without reprogramming the route. This repeat strategy worked for three pilots but it worked by chance, since enough time and distance had elapsed for the FMS to automatically advance to the next waypoint and thus for the route to be corrected.

None of the seven pilots who prematurely engaged the LNAV mode was able to explain the cause of the unexpected behavior prior to beginning recovery actions. The debriefing confirmed that the seven pilots who did not understand the observed LNAV behavior were either unaware of which waypoint was active during the event and/or were generally unfamiliar with floating waypoints and their effect on arming the LNAV mode after a deviation. One pilot believed that the unexpected LNAV behavior was a "malfunction."

VNAV ALT Mode

The canonical path for handling this event involves two steps: 1) entering the new cruise altitude into the FMS, and 2) pushing the altitude knob to make the FMS accept the new value (Figure 2). Completing these actions results in the activation of the VNAV PTH mode, which is necessary to achieve the desired descent profile. Otherwise, the automation remains in the VNAV ALT mode.

The event was "successful" in the sense that the VNAV ALT mode became active during cruise in ten of the twelve cases. In the other two cases, the pilots (4 and 10) proactively reprogrammed the FMS prior to reaching the new cruise altitude and went directly to the VNAV PTH mode. Three of the ten 'VNAV ALT' pilots (1, 2, and 3) successfully returned to the VNAV PTH mode by completing the canonical path.

The solution path for Pilot 3 is an example of a pilot who recovered the VNAV PTH mode from the VNAV ALT mode, though using an extraneous sequence of actions in addition to the canonical path . This strategy was described later by the pilot as "pushing buttons until it worked" and "resetting" the system, but also reflected incomplete knowledge of how to deal with this problem efficiently. Although the pilot did not understand why he was in the wrong mode, he knew that it was incorrect, and worked to resolve that discrepancy.

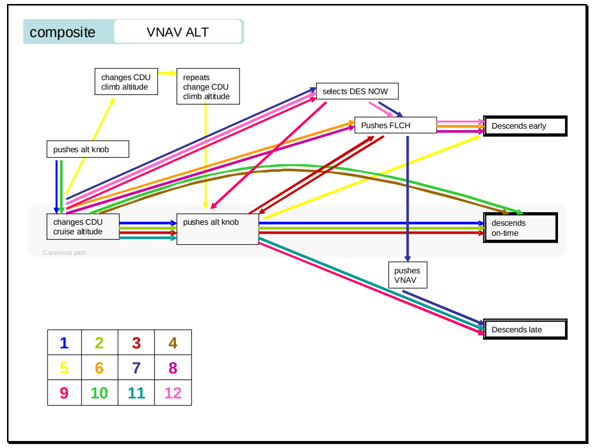


Figure 2. *Composite of abstracted solution paths for VNAV ALT event. Each line represents one pilot and is color-coded for all pilots.*

The other seven pilots remained in the incorrect mode (VNAV ALT) for a majority of the cruise phase. Note that there were no observable consequences of being in the VNAV ALT mode during this phase, since the aircraft was flying at a level altitude. Unwanted consequences would only appear when the aircraft reached the TOD point, approximately 20 minutes later, and would fail to begin the descent, creating the potential for the aircraft to miss programmed altitude restrictions. In the debriefing, all of these pilots were found to have gaps in their knowledge related to the functioning of the VNAV PTH and VNAV ALT modes. Interestingly, four of these seven pilots (5, 6, 8, and 12) avoided the consequence of the incorrect automation setting – the failure to descend automatically at the TOD – by deciding to descend earlier than the TOD point. In other words, the gaps in their mental model of the VNAV mode were either masked or worked-around by their early descent strategy, which they stated was based on the desire to alleviate workload during the descent. In contrast, three pilots (7, 9, and 11) remained in VNAV ALT at the TOD, and the aircraft did not descend as they had intended.

Of the three pilots (7, 9, and 11) who did not initiate an early descent, two recognized quickly that the aircraft had not started to descend and recovered by engaging the FLCH mode. One pilot (Pilot 11) was distracted with arrival preparations for almost 10 minutes after passing the TOD point, and recovered late by engaging the FLCH mode. During the event, none of these three pilots were able to explain why the aircraft did not descend as expected, suggesting an incomplete understanding of the automation that was later confirmed during the debriefing.

Discussion

All pilots completed the scenario “successfully” in the sense that they managed to complete the flight and land safely. At the same time, all participants experienced at least one disturbance during the course of the scenario. Note that these disturbances did not result from system faults. Rather, potentially unproblematic events were “managed” into disturbances from which pilots then had to recover.

One important goal of the current study was to explore the need for, and the effectiveness of, diagnosing errors and disturbances in the context of dynamic event-driven systems. In the present study, pilots rarely attempted to diagnose the source of a disturbance, except in two unsuccessful cases (two different pilots during two different events) in which pilots remained fixated on an incorrect diagnosis. This finding is in agreement with earlier findings from other dynamic domains where the absence of diagnostic activities was explained by time pressure and the need for immediate recovery to avoid negative consequences (Kanse and Schaaf, 2001; Kontogiannis, 1999; Reason, 1990). While time pressure and the immediate need to recover from disturbances (i.e., in cases of impending or actual deviations from assigned routes or altitudes) may have precluded diagnosis in many cases, it was also absent from contexts that were not time critical (i.e., the majority of the cruise phase in the VNAV ALT event). This may, in part, be explained by considerable knowledge gaps in pilot mental models of the automation which were observed in earlier research (Sarter and Woods, 2000; Mumaw et al., 2000) and confirmed in this study. For example, nine of 12 pilots in this study were found to have incomplete or inaccurate knowledge of the vertical navigation (VNAV) submodes of the FMS. These misconceptions – which were sometimes masked by serendipitous pilot actions that produced apparently seamless performance - likely contributed to problems with detecting, diagnosing and recovering from disturbances, and in some cases, even exacerbated the existing disturbance.

While the absence of diagnostic activities did not result in catastrophic outcomes, it may have affected the success and efficiency of recovery. In most cases, pilots used generic recovery strategies (repeating actions or resetting the automation) or engaged in trial-and-error behavior, rather than developing and implementing a problem-specific solution. In most cases, these generic recovery strategies, and also the observed tendency to use high levels of automation to manage disturbances (contrary to what is typically prescribed by training), were not successful and instead led to a delay in recovery, which further exacerbated the disturbance.

After detecting the disturbance in the LNAV Capture Event, pilots commonly resorted first to a “quick fix” by reverting to a lower-level mode (HDG SEL) in order to immediately correct the heading. This choice was likely prompted by the urgency of this disturbance which, over time, was producing an escalating divergence between the required and actual course. The use of such quick-fixes has been observed by other authors (Kontogiannis, 1999; Kanse and Schaaf, 2001) in process control domains. In those cases, they served to stabilize a situation in order to allow for an analysis of the problem and/or more thorough corrective actions. In contrast, five pilots in our scenario followed the “quick fix” with just a generic repetition of the LNAV engagement, without any further analysis or modification of the automation’s instructions.

The repetition strategy - observed primarily in the LNAV event - seemed to be based on pilots’ erroneous belief that the original action was appropriate but that the automation, for some reason, did not accept the pilot’s input or execute the command as intended. This example illustrates that coincidentally successful strategies can lead to erroneous beliefs which can become incorporated into a pilot’s mental model of the system. As a result, pilots may develop misrepresentations of functional system architecture that can lead to miscalibration of their system knowledge.

The resetting strategy - observed for 2 pilots during the VNAV ALT event - appears to be a type of workaround that did not require deeper system knowledge of how the disturbance occurred or how to avoid it in the future. Interestingly, both repetition and resetting strategies were observed by Plat and Amalberti (2000) in a simulator study of pilot responses to experimenter-induced software “bugs” or malfunctions in the behavior of the flight deck automation. This suggests that some pilots in our scenario treated disturbances as if they were discrete malfunctions which were unavoidable (i.e., not attributable to their actions) and required only generic fixes that did not require accurate or detailed system knowledge. However, these strategies can be brittle in that they may work in some contexts, but may not be effective in others, especially in unforgiving environments.

In addition, our findings indicate that disturbance management was not always well-supported by the available feedback to pilots. In the case of the VNAV ALT event, pilots were unable to visualize the implications of the active mode for the descent since there is no predictive vertical profile display. Instead, pilots are shown only a symbol and adjacent alphanumeric label ("T/D") representing the top-of-descent point on the map display. Aside from an alphanumeric mode annunciation on the PFD (i.e., "VNAV ALT") they receive no salient indication on the map display of whether the top-of-descent will be honored by the system. As a result, the current feedback may contribute to delays in detecting the error, which in turn, allow the disturbance in the aircraft's profile to escalate.

Conclusion

Error prevention alone will never be sufficient for improving safety in complex high-risk systems. Rather, a deeper understanding is needed of how human operators cope with the consequence of inevitable errors and thus the disturbances to the processes they monitor and control. The problems of inadequate feedback of autonomous system changes have been widely discussed (Sarter and Woods, 1995; Wiener, 1989) and have also been observed in the current study. However, these problems have often been discussed in the context of detecting the existence of an erroneous setting (e.g., "mode awareness" and "automation surprises"). Observations of pilot performance in the present study have shown that current automation design not only delays detection, but is too ambiguous for diagnosis, and does not support operators in recovering from disturbances in the most optimal way. Continued efforts in this area will inform the design of cognitive tools that effectively support this process.

Acknowledgements

The authors would like to thank Jack Bard, Randy Mumaw, Jack Butcher, Steve Kimball, and the airlines and pilots that participated in the study. This work was funded by a grant from the Federal Aviation Administration (Grant # **DTFA 96-G-043;** Technical monitor: Dr. Tom McCloy).

References

Boeing Commercial Airplane Group. (1994). *Statistical summary of commercial jet aircraft accidents: Worldwide operations 1959-1993* (Boeing Airplane Safety Engineering B-210B). Seattle, WA: Boeing.

Kanse, L. and Schaaf, T. (2001). Recovery from Failures in the Chemical Process Industry. *International Journal of Cognitive Ergonomics, 5*(3), 199-211.

Klinect, J.R., Wilhelm, J.A., & Helmreich, R.L. (1999) Threat and error management: Data from line operations safety audits. In Jensen, R. S. (Ed.), *Proceedings of the Tenth International Symposium on Aviation Psychology*, Columbus, OH: Ohio State University.

Kontogiannis, T. (1999). User strategies in recovering from errors in man–machine systems. *Safety Science, 32*, 49-68.

Mumaw, R.J., Sarter, N., Wickens, C., Kimball, S.,Nikolic, M., Marsh, R., Xu, W., & Xu, X. (2000). *Analysis of pilots' monitoring and performance on highly automated flight decks*. (Final Project Report: NASA Ames Contract NAS2-99074). Seattle, WA: Boeing Commercial Aviation.

Plat, M., & Amalberti R. (2000). Experimental crew training to deal with automation surprises. In N.B. Sarter and R. Amalberti (Eds.), *Cognitive Engineering in the Aviation Domain* (pp. 287-308). Hillsdale, NJ: LEA.

Reason, J. (1990). *Human Error*. New York: Cambridge University Press.

Sarter, N.B. and Woods, D.D. (1995). "How in the world did we ever get into that mode?" Mode Error and Awareness in Supervisory Control. *Human Factors, 37(1),* 5-19.

Sarter, N.B. and Woods, D.D. (2000). Teamplay with a Powerful and Independent Agent: A Full-Mission Simulation Study. *Human Factors*, *42*(3), 390-402.

Wiener, E.L. (1989). Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft. *NASA Contractor Report 177528*. Moffett Field, CA: NASA Ames Research Center.

Woods, D.D. (1984). Some results on operator performance in emergency events. In D. Whitfield, editor, *Ergonomic Problems in Process Operations*, Inst. Chem. Eng. Symp. Ser. 90, 21-31.

Woods, D.D. (1988). Coping with complexity: The psychology of human behavior in complex systems. In: Goodstein, L. P., Andersen, H. B., and Olsen, S. E. (Eds.), *Tasks, Errors, and Mental Models.* New York: Taylor & Francis.

Woods, D.D. (1993). Process tracing methods for the study of cognition outside of the experimental psychology laboratory. In G. Kln, J. Orasanu, & R. Calderwood (Eds.), *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex Publishing Corporation.

Woods, D.D. (1994). Cognitive Demands and Activities in Dynamic Fault Management: Abduction and Disturbance Management. In N. Stanton, editor, *Human Factors of Alarm Design*, Taylor & Francis, London, 1994.

Woods, D.D., Johannesen, L.J., Cook, R.I., & Sarter, N.B. (1994). *Behind human error: Cognitive systems, computers, and hindsight.* WPAFB: CSERIAC.

MISHAP TRENDS IN SEVERAL AIR FORCE AIRCRAFT: IMPLICATIONS FOR CRM TRAINING

Robert Nullmeyer
Air Force Research Laboratory
6030 S. Kent Street
Mesa Arizona
Bob.nullmeyer@mesa.afmc.af.mil
(480) 988-6561, ext. 283

Lt Col David Stella
Air Force Research Laboratory
6030 S. Kent Street
Mesa Arizona
David.stella@mesa.afmc.af.mil
(480) 988-6561

Gregg Montijo
Crew Training International
1811 North Fourtyniner Drive
Tucson Arizona 85749
gmontijo@cti-crm.com
(520) 240-0827

Human factors trends in C-130, F-16, and A-10 mishaps were reviewed for relevance to cockpit/crew resource management (CRM) course content. The current Air Force Safety Center human factors taxonomy includes about 360 detailed human factors elements. About sixty of these taxonomy elements map directly into the six CRM core areas identified in Air Force Flying Operations publications (communication, risk management/decision making, situational awareness, task management, crew coordination/flight integrity, and mission preparation/ debriefing). This small fraction of human factors elements accounted for well over half of the causal and strongly contributing factors cited in each platform. The relative contributions of specific CRM core areas varied across applications. Tactical airlift mishap CRM factors were fairly uniformly distributed across all six traditional CRM areas. In F-16 and A-10 mishaps, task management and situational awareness were particularly frequent causal and major contributing factors. Planning, flight integrity, and communication were rarely cited. We describe the mishap data that are available from the Air Force Safety Center, our analytic approach, trends identified, and implications for CRM training. We anticipate that these analyses will contribute to better focused CRM training objectives and course content that will, in turn, enable CRM training to be a major contributor to the success of recent Department of Defense efforts to reduce preventable mishaps.

Introduction

Secretary Rumsfeld challenged the Services to reduce mishap and accident rates by at least 50% over a two year period. A Joint Service Safety Conference (JSSC) was established to develop a unified approach for meeting the Secretary's challenge. Many researchers have documented the large role played by human factors in flight mishaps. For example, Helmreich and Fouchee (1995) reported that flight crew actions were causal in more than 70% of worldwide accidents from 1959 to 1989 involving aircraft damage beyond economic repair. Similarly, Luna (2001) reported that human factors were major contributors or causal in over 60% of Air Force Class A mishaps from 1991 to 2000. Such long term trends suggest that meeting the Secretary's challenge will require solutions to human factors problems, and as a result, a Human Factors Working Group was established as a critical part of the JSSC. Analyses of recent aviation mishaps across the services by this working group revealed that Crew Resource Management (CRM) issues are still frequently cited in aviation mishap reports across the services.

Helmreich, Klinect and Wilhelm (1999) define CRM skills as "a primary line of defense against the threats to safety that abound in the aviation system and against human error and its consequences" and state that, to be effective, CRM training must be based on detailed knowledge of current safety issues. CRM training requirements for Air Force aviators reflect a similar safety focus in Air Force Instruction (AFI) 11-290, *Cockpit/Crew Resource Management Training Program* (2003). AFI 11-290 states that the objective of CRM training is to "develop aircrew skills in recognizing and responding to the conditions that lead to aircrew error." Six core curriculum areas are specified: situational awareness (SA), risk management/decision making, mission planning/ debrief, task management, crew communication, and coordination/flight integrity.

Helmreich, et al. (1999) identified five critical data sources: 1) formal evaluations of flight crews; 2) incident reports; 3) surveys of flightcrew perceptions regarding safety and human factors; 4) information on parameters of flight from flight data recorders; and 5) line operations safety audits (LOSA). Each illuminates a different aspect of flight operations. They proceeded to explore lessons learned from LOSA data.

Given the numbers of human factors-related Class A mishaps (loss of life, a destroyed air frame, or more that $1 mission damage), it only makes sense to learn as much as possible about the factors that most often led to these outcomes in the past. Mishap summaries are often used to develop case studies for CRM training and guide content of simulator refresher scenarios.

The full Class A mishap reports also include much more detailed descriptions of the human factors that caused or contributed to the undesired outcome. Unfortunately, analyses and application of these detailed human factors data have been rare in the training community. That picture is changing. CRM factors in C-130 Class A mishaps were recently analyzed (Nullmeyer, Stella, Flournoy, and White, 2003) as part of a larger program to improve CRM instruction for C-130 tactical airlift crews. Elements from all six core CRM areas were frequently cited in C-130 mishaps from 1990 through today. Within each CRM area, however, a small subset of elements accounted for the vast majority of causal or strongly contributing factors. This information was used to focus C-130 CRM training content on particularly problematic elements (Deen and Wilson, 2003).

Based on this initial success, analyses were recently expanded to include A-10 and F-16 Class A mishaps. Our focus in this paper is on major trends found in the more detailed C-130, F-16, and A-10 mishap reports, including commonalities and differences across platforms. We recognize that mishap reports are not sufficient by themselves to structure CRM training. Maurino (1999) correctly states that if we only look at accidents and incidents, we only learn about CRM failures. Dekker (2003) describes several potential problems with over-reliance on human error taxonomies, including risks associated with removing the context that helped produce the error.

These concerns suggest that detailed mishap human factors trends need to be viewed in the context of other information to develop truly robust CRM training. For example, instructor comments in student records were reviewed and CRM behaviors exhibited in annual simulator training were captured as part of the earlier review of C-130 CRM training (Spiker, Wilson and Deen 2003). Both enabled visibility into both positive and negative behaviors, and the simulator study in particular, allowed naturalistic observations of crew interactions and mission performance in the context of complex and demanding simulator scenarios.

Mishap Data Sources

The Air Force Safety Center documents Class A mishaps at varying levels of granularity. The analyses reported here combine data from all four data sources.

The Air Force Safety Center home page (http://afsafety.af.mil/) provides considerable summary mishap statistical information including hours flown and mishap frequencies by aircraft type, by year. Mishap frequency counts were used to check the completeness of other data sources. Flying hours per year were essential for determining mishap rates per 100,000 flying hours.

Mishap Summaries are executive summaries of the Safety Investigation Board's report (Tab T of the full report). They include information such as the mishap date, location, day or night, type of mishap (e.g, midair collision), phase of flight, and other descriptive data. It provides a one paragraph description of the mishap, and lists findings and recommendations.

A detailed **Human Factors Database** is populated and maintained by Air Force Safety Center Life Sciences analysts who use a common human factors taxonomy to structure findings regarding role played by operators, maintainers, and other personnel in each Class A mishap. The database includes dozens of fields. In the analyses reported here we focused on the human factors that were cited along with a rating for each factor ranging from "causal" (4) major contributor (3) and minor contributor (2), to minimal contributor (1) that indicates the degree to which each factor was involved in the outcome.

A **Life Sciences Report** is part of the full Class A mishap report (Tab Y). It provides fairly detailed discussions of **each** element cited in the human factors database and identifies interrelationships among the human factors. These discussions are extremely useful for understanding the actual behaviors underlying the human factors data base entries.

Time Frames of Anaylses. Mishap frequencies by aircraft type and year were used to determine the time periods to be included in subsequent analyses. As can be seen in Figure 1, there have been many more F-16 mishaps in the past few years than A-10 or C-130 mishaps. In an effort to achieve a reasonable sample size and maximize currency, we analyzed F-16 mishaps from 2000 through 2004, but expanded the time frame back to 1995 for C-130 and A-10 mishaps. These time frames resulted in 31 F-16 mishaps, 20 A-10 mishaps, and 8 C-130 mishaps.

Data Structure. The Life Sciences Branch, Aviation Safety Division of the Air Force Safety Center provided access to A-10, F-16, and C-130 databases to identify human factors that caused or contributed to Class A mishaps in these Air Force communities. The Air Force Safety Center's human factors taxonomy was first reviewed to identify elements that are relevant to CRM. About sixty of the 360 detailed taxonomy elements were determined to be CRM-related. These were then mapped into to the six CRM areas specified in Air Force Instruction (AFI) 11-290 as follows:

Perceptual and attention management elements were mapped into situational awareness (SA).

Task management factors included procedural elements and task misprioritization.

Risk management and decision making elements came primarily from the judgment and decision making node of the mishap taxonomy.

In-flight analysis and in-flight planning were added to preparation factors to create the mission planning and debriefing.

Communication was a preexisting node in the mishap taxonomy that encompassed both intra-cockpit interactions and interactions with external to the aircraft.

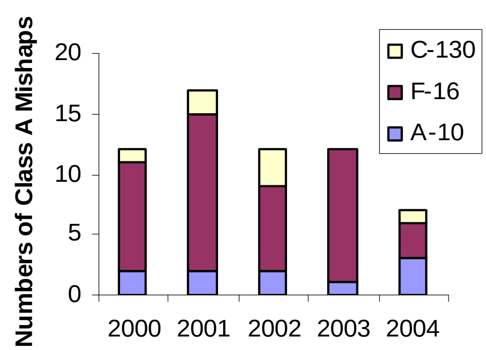
Elements of the cockpit/ crew resource management node (e.g., leadership, subordinate style and crew coordination) were combined with hazardous attitude elements based in the definition of crew coordination/flight integrity provided in AFI 11-290.

Results

The numbers of mishaps in which each CRM area was represented as least once as a causal or strongly contributing factor is depicted in Figure 2 as an annualized number. In the past decade (1995-2004) , there were over 100 F-16 Class a mishaps, 19 A-10 Class A mishaps, and 8 C-130 Class A mishaps. We included all of these A-10 and C-130 mishaps in this analysis. Due to the large numbers of F-16 mishaps, we focused on mishaps from the last five years. Twenty one of these mishaps were attributed to human factors. The remaining Class A mishaps were primarily loss of engine or bird strike, for which human factors were not cited.

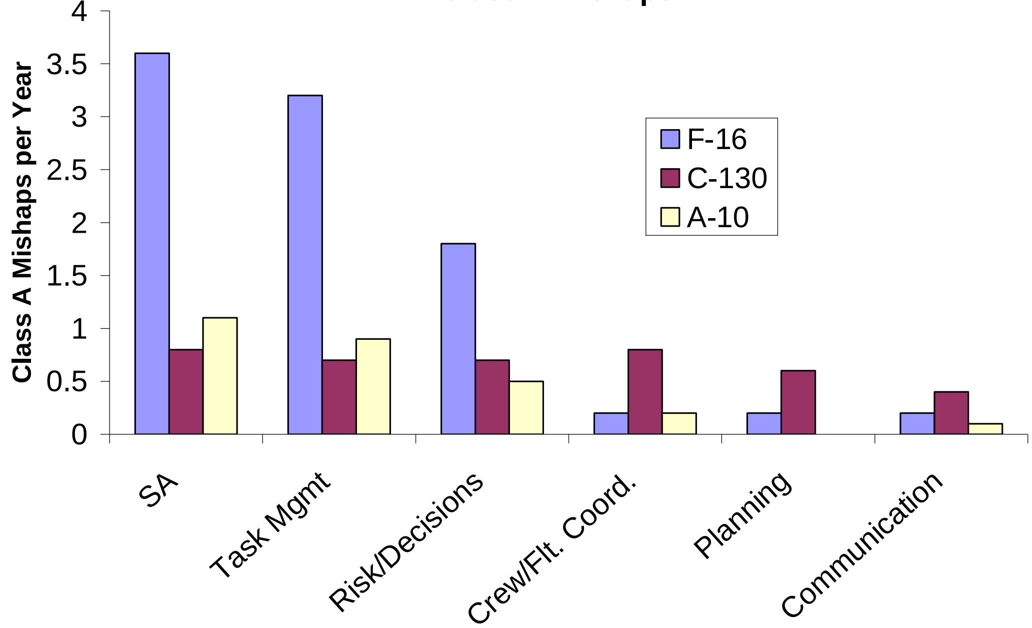
Mishap frequencies. The numbers of Class A mishaps over the past 5 years are shown in Figure 1 for C-130, A-10, and F-16 aircraft. There were notably few F-16 Class A mishaps in the most recent year (2004).

Figure 1: Class A Mishap Frequencies 2000-2004



CRM as a Causal or Major Contributing Factor. CRM-related factors and their numerical ratings were extracted from the human factors database for each mishap. Data from individual CRM-related factors were combined into the six CRM dimensions specified in AFI 11-290. From these consolidated data sets, we determined the number of mishaps in which a CRM dimension was cited at least once as a causal or major contributing factor. The resulting frequencies

Figure 2: CRM Core Areas as Causal or Major Factors in Class A Mishaps



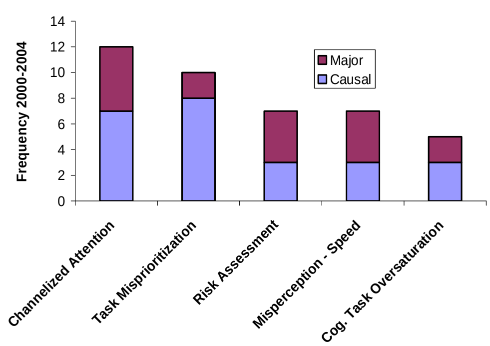
were converted into frequencies per year. The resulting rates are depicted in Figure 2. In C-130 Class A mishaps, causal and major contributing factor rates were fairly evenly distributed across the six CRM dimensions. For A-10 and F-16 mishaps, however, rates were much higher in some CRM areas than others. Rates were particularly uneven For F-16 mishaps. SA, task management, and risk management factors were cited frequently. Planning, flight integrity and communication were rarely cited.

Underlying CRM-Related Factors **-** We now shift the focus to the specific CRM–related human factors that were most frequently cited as being causal or strongly contributing in Class A mishaps. The top five factors are first listed for each platform. Commonalities and differences across platforms are then discussed.

F-16 CRM-Related Factors. The five specific human factors that were most frequently cited in F-16 Class A mishaps from 2000 through 2004 are shown in Figure 3. The first, fourth and fifth most frequent F-16 factors were directly related to SA. The remaining two were directly related to task management and risk assessment. In fact, all 10 leading human factors in F-16 mishaps were related to the SA, task management, or risk assessment/decision making areas of CRM

.Channelized Attention, cited most frequently, is a factor when the pilot is focusing conscious attention on a limited number of environmental cues to the exclusion of others of subjectively equal, higher or more immediate priority leading to an unsafe situation. Recent examples included attending to broken equipment inside the cockpit during low level flight, and relying exclusively on the Radar/Electro-Optical (REO) display while ignoring all other instruments, resulting in a failure to recognize the distance to the runway and altitude relative to the rising terrain.

Figure 3: Most Frequent Factors in F-16 Class A Mishaps



Task Misprioritization is a factor when the individual does not organize, based on accepted prioritization techniques, the tasks needed to manage the immediate situation as perceived by the individual.

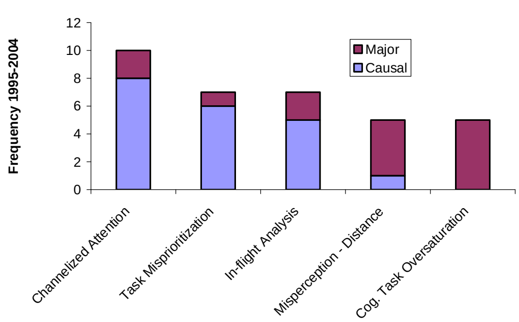
Risk Assessment is a factor when the individual fails to adequately evaluate potential risks associated with a selected course of action and this failure leads to an unsafe situation. Behaviors labeled *risk assessment* varied greatly across accidents.

Cognitive task oversaturation occurs when the quantity of information to process exceeds a person's cognitive or mental resources, resulting in a loss of SA.

Specific A-10 CRM Factors. The five leading human factors cited in A-1- mishaps are depicted in Figure 4. The top two were directly related to the CRM areas of dimensions of SA and task management. Definitions were discussed in the previous section. The third element, in-flight analysis, refers to a failure to analyze an in-flight situation to the extent normally expected which leads to degraded performance. This factor was assigned to the *decision making* area of CRM in our quantitative analyses. Misperceived distance and cognitive task oversaturation round out the top five. Both factors were described in the preceding section. Consistent with the overall CRM patterns in A-10 mishaps discussed earlier, these specific underlying CRM-related factors reflect problems with SA, task management, and decision making.

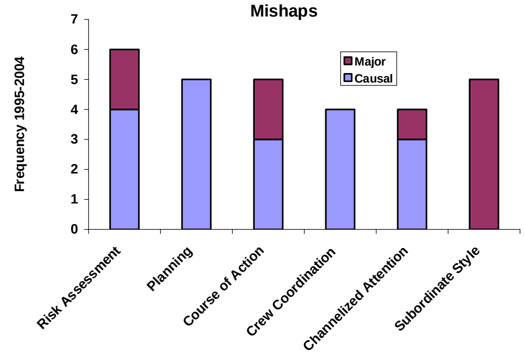
Two hazardous attitudes, overconfidence and complacency, were in the top ten factors. AFI 11-290 places such factors under crew coordination/flight integrity.

Figure 4: Most Frequent CRM Elements in A-10 Class A Mishaps



Specific C-130 CRM Factors. The top individual CRM-related factors in C-130 mishaps from 1995 through 2004 are summarized in Figure 5. There was a two-way tie for the fifth factor between channelized attention and subordinate style--both are presented and discussed.

Figure 5: Most Frequent CRM factors in C-130 Class A Mishaps



Two factors, risk assessment and channelized attention were discussed in earlier sections. *Flight planning* is a factor when proper flight planning for the mission is not accomplished. In most of these mishaps, other military duties competed with planning activities, resulting in the crew failing to access accessing available, crucial information. *Course of action selected* is a factor when the wrong course of action is selected through faulty logic and decision making. Several instances originated in planning due to inadequate gathering of data that was readily available (e.g., terrain, weight of cargo, or weather). Other instances involved less-than-ideal responses to in-flight equipment problems.

Crew Coordination is defined as the lack of a systematic division of subtasks between crew of flight members to accomplish a larger task more efficiently. Behaviors leading to this factor being cited included lack of cross-check, failure to provide feedback, lack of input, not catching checklists that were started but not completed, failure to delegate backup responsibilities, and lack of a symmetrical division of tasks within the cockpit.

Subordinate style/copilot syndrome refers to the basic belief by an aviator that someone else (other crewmembers or individuals external to the aircraft) have the situation under control and are looking out for their best interest. Several instances involved a well respected individual on the crew with whom others felt they did not need to be directive, resulting in some crewmembers taking themselves out of the decision process. Other mishap involved misplaced trust in planners or air traffic control.

The CRM-related causal or contributing factors cited in C-130 mishaps were consistent with the AFI 11-290 set of six core CRM areas. Four of the six areas were included in the top six factors. All six were represented in the top ten with the inclusion of intracockpit communication and necessary action delayed (a task management factor).

Commonalities and differences across platforms

Human factors remained prominent in recent F-16, A-10, and C-130 Class A mishaps. Further, the most frequency cited human factors were consistently CRM-related. As a result, CRM skills remain great targets of opportunity for reducing preventable mishaps in all three platforms. SA, task management, and risk management/decision making factors were evident across all three air frames.

The relative contributions of the remaining core CRM areas, however, appeared to differ substantially across air frames. Human factors related to *crew coordination/flight integrity* were more common and central in C-130 mishaps than were factors in any other CRM area. The problems revolved around interpersonal interactions--failure to back up other crew members or question an unsafe condition or action. The crew coordination/flight integrity factors for both F-16 and A-10 are limited to hazardous attitudes—overconfidence, complacency, invulnerability, and get-home-it is.

<u>Mission planning</u> was causal in the majority of C-130 mishaps included in this analysis, yet pre-mission planning was never cited in either F-16 or A-10 mishaps. The small presence of mission planning in F-16 mishaps came from a single instance of faulty in-flight replanning.

<u>Communication</u> was the least frequently cited CRM area in all three air frames. Intracockpit communication was the leading C-130 communication problem. Misinterpreted communication and external communication. Were occasional problems in all three aircraft.

Conclusions

First and foremost, the most frequently cited causal and major contributing factors to flight mishaps in the mishap reports that we reviewed were consistently CRM-related. The six core CRM areas in AFI 11-290 are broad enough to cover at least the most frequently cited factors. In single seat aircraft, some CRM areas do not appear to be as problematic as others. Specifically, mission planning, communication, and flight integrity are seldom cited as causal or major contributing factors in A-10 and F-16 mishaps. The remaining core CRM areas (SA, task management and risk assessment/decision making are areas that will need to be improved if mishaps are to be reduced. Even within these core CRM areas, the majority or problems are clustered in a few factors. As a result, we can be very prescriptive concerning areas in which improvement should impact mishap rates.

The bottom line is that AFI 11-290 defines a sound domain for CRM training, but our data suggest that, at least for single seat aircraft, attending to a few particularly troublesome areas could pay big dividends.

References

Dekker, S.W.A. (2003). Illusions of explanation: A critical essay on error classification. *The international Journal of Aviation Psychology, 13,* 2, pages 95-106.

Deen, G.C., & Wilson, D.D. (2003). Using multiple sources to upgrade a successful CRM program. In *Proceedings of the International Symposium of Aviation Psychology, 2003.* Dayton, OH.

Helmreich, R.L. & Foushee, H.C. (1993). Why crew resource management? Empirical and theoretical bases of human factors training in aviation. In E.L. Wiener, B.G. Kanki, and R.L. Helmreich, *Cockpit resource management*, San Diego, CA: Academic Press.

Helmreich, R.L., Klinect, J.R., & Wilhelm, J.A. (1999). Models of threat, error, and CRM in flight operations. In *Proceedings of the Tenth International Symposium on Aviation Psychology* (pp. 677-682). Columbus, OH.

Luna, T.D. (2001). USAF aviation safety: FY 2000 in review. Available at http://safety.kirtland.af. Mil.

Maurino, D.E. (1999). Safety prejudices, training practices, and CRM: A midpoint perspective. *The International Journal of Aviation Psychology, 9,* 413-422.

Nullmeyer, R., Stella, D., Flournoy, J. & White, D. (2003). Using Air Force mishap data to improve C-130 CRM training. In *Proceedings of the International Symposium on Aviation Psychology*, Dayton, OH.

Spiker, V.A., Wilson, D.D., & Dean, G.C. (2003). CRM and mission performance during C-130 mission oriented simulator training. In *Proceedings of the International Symposium on Aviation Psychology*, Dayton, OH.

A JOINT SERVICE DESCRIPTION OF CREW RESOURCE MANAGEMENT FOR ACCIDENT ANALYSIS AND PREVENTION

Robert Nullmeyer
Air Force Research Laboratory
6030 S. Kent Street
Mesa Arizona 85281
Bob.nullmeyer@mesa.afmc.af.mil
(480) 988-6561, ext. 283

LCDR Deborah White
Naval Safety Center
375 A Street
Norfolk, Virginia 23511
Deborah.j.white@navy.mil
(757) 444-3520, ext. 7231

A Department of Defense (DoD) Cockpit/Crew Resource Management (CRM) working group was established to develop a common definition of CRM and create a method for capturing CRM factors in mishap investigations to be used by the Army, Navy, Air Force and Coast Guard investigators and analysts. The goal of the CRM working group was to build on the substantial commonalities among the services regarding CRM concepts, while also accommodating service-unique requirements. This CRM project was part of a larger DoD initiative to promote a common Human Factors taxonomy, investigation, and analysis system for DoD-wide implementation. Shappell and Weigmann's Human Factors Analysis and Classification System (HFACS) formed the structure for both the CRM working group and the larger initiative to standardize human factors and human performance terminology across the DoD. A common CRM definition was developed and specific codes were created that can be translated into Army, Navy, Coast Guard, or Air Force CRM skill sets. The resulting specific codes are located in several areas of the HFACS taxonomy. Given the success of the HFACS project, the CRM working group is now exploring ways to further standardize CRM concepts and improve CRM training across services.

Introduction

In response to rising mishap rates across the military Services, Department of Defense (DoD) Secretary Rumsfeld issued a challenge to reduce preventable mishaps by at least 50% in two years (Rumsfeld, 2003). Over 60% of all Air Force Class A military aviation mishaps and over 90% of fatalities in the past decade were attributed to human factors (Luna, 2001). Similar patterns are found in the other services. As a result, the role of human factors in mishaps must be addressed if we are to meet the Secretary's challenge.

The Army, Navy, Marine Corps, Air Force, and Coast Guard established a joint Human Factors Working Group to develop a unified set of solutions for these human factors problems. One initiative of this group is to develop a human factors mishap taxonomy that is accepted and used by all DoD components to enhance the ability to share accident information among the services. The Human Factors Analysis and Classification System (HFACS) that was created by Shappell and Wiegmann (2000) was adopted as the basic structure, with the understanding that some tailoring might be appropriate to meet the needs of the DoD and each individual military service.

The HFACS model describes four levels of human failure, each influencing the next. Working backward from the mishap, the first level of the DoD version of HFACS depicts the **acts** of operators that ultimately led to the mishap. The next level involves the condition of the aircrew as it affects the performance of the crew. This level is refered to as **preconditions** and includes conditions such as fatigue, perception and cognition. Poor communication and coordination practices comprise *Crew Resource Management* (CRM), which is a node in this level. The remaining levels address **supervision** and **organizational influences**.

CRM was identified as a potential stumbling point. A lack of consensus concerning its definition and scope is well documented in the scientific literature. Helmreich, Merritt, and Wilhelm (1999) identified five specific generations of CRM and reported that all are in use somewhere today.

Within the military, substantial discrepancies in the meaning of this term can also exist among mishap investigators, researchers, and trainers. For example, Salas, Prince, Bowers, Stout, Oser, and Cannon-Bowers (1999) documented considerable divergence in aviation training concerning both the basic definition of CRM and the domain that is covered. They went on to recommend that focusing CRM for Naval aviators on "a set of teamwork competencies that allow the crew to cope with situational demands that would overwhelm any individual crewmember" based in part on analyses of commercial air carrier accident reports. This definition of CRM was also used in a recent analysis of military mishaps (Wilson-Donnelly and Shappell, 2003).

Nullmeyer, Stella, and Montijo (2005) compared causal and major contributing factors across military transport, fighter, and attack aircraft mishaps and found that teamwork factors were frequently cited in transport mishaps, but were rarely cited in single seat aircraft mishaps, where situational awareness and task management issues were more prominent.

A joint service CRM working group was established to develop a common CRM vision for the DoD safety community that still accommodates the needs of other parts of the organization, and then incorporate this shared vision into the emerging DoD HFACS structure. CRM training program managers, safety center analysts, instructors and researchers from all military services, the Coast Guard and industry participated. These stakeholders found considerable overlap among high-level service descriptions and a core definition was developed, focusing on the effective use of all information by individuals or crews.

Currently, each service has a somewhat unique list of underlying CRM elements. The CRM working group reviewed the emerging DoD HFACS taxonomy, the joint CRM definition, each Service's underlying elements, and trends identified in several recent military mishap meta-analyses. The basic HFACS structure accommodated the joint view of CRM well. Several elements of the DoD HFACS, however, seemed to warrant reconsideration. The *Crew Resource Management* node of HFACS addressed interpersonal interactions during planning and execution. This represents a sensible bundling of related behaviors, but all service CRM skill sets go well beyond this to include situation awareness, decision making and other issues. As a result, the meaning of *Crew Resource Management* in the mishap taxonomy did not match the meaning attributed to the term by operators and trainers. Perhaps ironically, the first conclusion reached by the CRM working group was that the term "*crew resource management*" needed to be removed from the HFACS structure. The goal was to avoid confusion arising from its multiple meanings. The team interaction node was kept but relabeled *Coordination, Comunication, and Planning* to more accurately reflect its content. Other CRM areas that are included in individual service CRM skill sets were also found in multiple parts of the HFACS structure such as cognitive factors, perceptual factors, and decision errors.

In this paper, we describe service-specific CRM definitions and scope, summarize commonalities and differences across the services, and distil a common working definition. We also describe our recommendations regarding CRM elements both within the category of *Coordination, Communication, and Planning* and in other areas of the new DoD HFACS. We believe that the resulting representation of CRM behaviors will result in better categorization of CRM problem areas by mishap investigators and in more useful search tools for use by analysts.

CRM in DoD Training

Aircrew Coordination Training (ACT) in the Army and the CRM training in the Navy, Coast Guard, and Air Force have highly similar desired outcomes. The objective of ACT is "to provide aircrews the knowledge, skills, and attitudes necessary to increase their mission effectiveness, while decreasing the errors that lead to accidents" (Katz and Grubb, 2003). Air Force and Navy objectives are indistinguishable from this. The Air Force Instruction that establishes CRM training requirements states that the goals are to maximize operational effectiveness and combat capability and to preserve Air Force personnel and material resources and states that the objective is to develop aircrew skills in recognizing and responding to the conditions that lead to aircrew error (Air Force, 2003).

Each Service provides its own dimensions to organize targeted ACT or CRM skills, as follows:

Navy—decision making, assertiveness, mission analysis, adaptability/flexibility, communication, leadership, and situational awareness.

Army—team relationships, mission planning and rehearsal, workload, exchange of mission information, and cross-monitoring performance.

Air Force— mission planning and debrief, flight integrity/crew coordination, situational awareness, risk assessment/decision making, communication, and task management.

There is clear overlap among these service-specific CRM dimensions to be addressed. There are also differences regarding both terminology and, to a lesser extent, scope. Of some note, each Service's targeted CRM skills go well beyond the HFACS elements originally listed under *Crew Resource Management:*

- Crew coordination/flight integrity
- Communication
- Mission preparation
- Analysis
- Mission in progress planning
- Crew leadership
- Authority Gradient

CRM Working Group Products

To develop the requested common CRM vision and produce a blueprint for incorporating this vision in the emerging DoD HFACS structure, the CRM working group first developed an overarching definition of CRM that reflected both definitions in the scientific literature and individual Service philosophies. Next, the original *CrewRresource Management* node was relabeled to focus on crew and team interactions, an area that was part of every Service's CRM concept. Each service had valuable elements to contribute to this node. These were combined to eliminate redundancy and ensure coverage. Finally, the larger HFACS was reviewed to ensure that other individual service CRM skills were addressed.

Three major products were developed: (1) a DoD wide CRM definition; (2) an HFACS node with subordinate codes to address crew and team interaction issues; and (3) a guide to other parts of the HFACS taxonomy that reflect other CRM behaviors.

A DoD CRM Definition. The effective use of all available resources by individuals, crews and teams to safely and efficiently accomplish the mission or task. [CRM training is a key component of a combined effort to identify and manage the conditions that lead to error.]

<u>Modified HFACS precondition category:</u> *Coordination/ Communication/ Planning* (replaces Crew Resource Management).

Coordination/Communication/Planning factors refer to interactions among individuals, crews, and teams involved with the preparation and execution of a mission that resulted in human error or an unsafe situation.

- **Crew/team leadership** is a factor when the crew/team leadership techniques failed to facilitate a proper crew climate, to include establishing and maintaining an accurate and shared understanding of the evolving mission and plan on the part of all crew or team members.

- **Cross-monitoring / backup** is a factor when crew or team members failed to monitor or back-up each other's actions and decisions.

- **Task delegation** is a factor when the crew or team members failed to actively manage the distribution of mission tasks to prevent the overloading of any crew member.

- **Rank/position authority gradient** is a factor when a crew or team member allowed differences in rank or experience to influence their willingness to speak up or actively interact with all members of the team.

- **Assertiveness** is a factor when individuals failed to state critical information or solutions with appropriate persistence.

- **Critical/accurate information communicated** is a factor when known critical information was not provided to appropriate individuals in an accurate or timely manner. This covers both inadequate intra-cockpit and external communication.

- **Standard/proper terminology** is a factor when clear and concise terms and phrases per service standards and training were not used.

- **Challenge and reply** is a factor when communications did not include supportive feedback or acknowledgement to ensure that personnel correctly understand announcements or directives.

- **Mission planning** is a factor when an individual, crew or team failed to complete all preparatory tasks associated with planning the mission, resulting in an unsafe situation. Planning tasks include information collection and analysis, coordinating activities within the crew or team and with appropriate external agencies, contingency planning, and risk assessment.

- **Mission briefing** is a factor when information and instructions provided to individuals, crews, or teams were insufficient, or participants failed to discuss contingencies and strategies to cope with contingencies.

- **Mission-in-progress re-planning** is a factor when crew or team members fail to adequately reassess changes in their dynamic environment during mission execution and change their mission plan accordingly to ensure adequate management of risk.

CRM Factors in Other HFACS Areas. Each service has invested considerable effort to integrate a consistent CRM concept throughout training and operational documents. For example, key CRM skill areas for Air Force aviators are articulated in an 11-series instruction, which makes it a flying operations publication. The core CRM skills are further amplified in aircrew evaluation criteria and training regulations. In each service, the scope of CRM in these training and operations documents goes well beyond coordination, communication, and planning.

In order to capture these broader CRM concepts, the working group reviewed the larger HFACS taxonomy and identified factors that are directly related to common CRM training areas. In a few instances, we recommended adding a new factor to an existing HFACS node to reflect an area that was previously not covered.

Crew/team coordination, communication, and mission preparation elements were addressed in the modified node discussed above. The main CRM elements distributed in other parts of the HFACS structure are shown in Table 1. CRM factors external to coordination communication, and planning addressed workload/task management, risk management/decision making, situational awareness, and hazardous attitudes. Each element is followed with the node in HFACS where it is located. Not all services include all of these elements in their CRM targeted behaviors at this time. Analysts can easily include or exclude individual elements as they see fit.

Within the larger HFACS structure, CRM elements tend to be clustered in either the *judgment and decision-making errors* node of "Acts" or in one of four preconditions—*perceptual errors, cognitive errors, pshco-behavioral factors, or coordination/ communication/planning factors.*

CRM Conclusions and Next Steps

Our working group focused on CRM (or ACT) as it is defined and taught in the military services and Coast Guard. We discovered substantial agreement on program goals, high level definitions, and even scope. We found less agreement on terminology, even at the highest level where some refer to the topic area as "ACT" while others called it "CRM".

Although we found considerable consistency across the services regarding the meaning of CRM or ACT in operations and training, the HFACS node labeled *Crew Resource Management* reflects a much narrower range of factors. One consequence of this difference is be that a key word search for CRM factors in mishap reports would mask numerous non-team interaction factors concerning the effective use of all resources, yielding potentially misleading results if the requestor does not comprehend the tighter focus of the HFACS category. Our short term solution was to avoid using the term and to focus instead on underlying behaviors.

Table 1. *Traditional CRM <u>Training</u> Areas in DOD HFACS v5.6 with CRM Working Group Factors*

Workload/Task management

Cognitive task oversaturation [Cognitive factors]
Task Misprioritization [Judgment & Decision-Making Errors]
Necessary Action – Rushed [Judgment & Decision-Making Errors]
Necessary Action – Delayed [Judgment & Decision-Making Errors]

Situational Awareness

Illusion – Kinesthetic [Perceptual Factors]
Illusion – Vestibular [Perceptual Factors]
Illusion – Visual [Perceptual Factors]
Misperception of flight Condition [Perceptual Factors]
Misinterpreted/Misread Instrument [Perceptual Factors]
Expectancy [Perceptual Factors]
Auditory Cues [Perceptual Factors]
Spatial Disorientation (Type 1) Unrecognized [Perceptual Factors]
Spatial Disorientation (Type 2) Recognized [Perceptual Factors]
Spatial Disorientation (Type 3) Incapacitating [Perceptual Factors]
Temporal Disorientation [Perceptual Factors]
Inattention [Cognitive Factors]
Channelized Attention [Cognitive Factors]
Confusion [Cognitive Factors]
Distraction [Cognitive Factors]
Geographic Misorientation (Lost) [Cognitive Factors]
Unaware of External Hazard [Cognitive Factors]
Unaware of Equipment System Status [Cognitive Factors]
Lack of Task Awareness [Cognitive Factors]
Complacency [Psycho-Behavioral Factors]

Decision making/risk assessment

Decision Making During Operations [Judgment & Decision-Making Errors]
Risk Assessment – Formal [Supervision--Planned Inappropriate Operations]
Risk Assessment – During Operations [Judgment & Decision-Making Errors]
Caution/Warning Ignored [Judgment & Decision-Making Errors]

Hazardous Attitudes

Overconfidence [Psycho-Behavioral Factors]
Motivation – Inadequate [Psycho-Behavioral Factors]
Motivation – Misplaced [Psycho-Behavioral Factors]
Motivation to Succeed – Excessive [Psycho-Behavioral Factors]
Get-Home-It is/Get-There-It is [Psycho-Behavioral Factors]
Motivational Exhaustion (Burnout) [Psycho-Behavioral Factors]

The working group merged each service's unique CRM skills into the DoD HFACS model. The resulting set of CRM codes represented a macro CRM skill list for the DoD, from which individual Service CRM skill areas can be easily constructed, and in fact, the macro list does not deviate very far from any service's functional scope for CRM training. From this experience, the CRM working group members are optimistic that the DoD HFACS model will facilitate sharing of information across services and still meet the needs of the participating organizations.

The original working group tasking was to develop standardized mishap investigation terminologies for CRM factors that will be used by all services. We believe the recommended changes to the DoD HFACS will do this. Given the clear similarities of underlying CRM elements that emerged across services, working group participants recognized an opportunity to expand cooperation beyond development of common HFACS codes. Each service currently has its own CRM training guidance, language system, and training strategies. Targeted CRM behaviors are organized into seven dimensions in Navy training, six in the Air Force, five in the Army, and four in the Coast Guard. As joint training and joint military operations become more common, these service-unique practices appear to be increasingly counter-productive.

Common sense suggests that a shared CRM language system is both possible and appropriate. In addition, the services have overlapping needs to better understand root causes, develop CRM training strategies, and assess the effectiveness of various approaches to training. Progress will clearly be maximized if the services can build on each other's advances.

References

AFI 11-290 (1998, July). *Cockpit/crew resource management program*. Washington, DC: Headquarters US Air Force.

Helmreich, R.L., Merritt, AC., & Wilhelm, J.A. (1999). The evolution of crew resource management training. *The InternationalJournalof Aviation Psychology*, *9*,(1), 19-32.

Katz, L.C. & Grubb, G.N. *Enhancing U.S. Army aircrew coordination training.* (ARI Special Report #56). Alexandria VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Luna, T.D. (2001). USAF aviation safety: FY 2000 in review. Available at http://safety.kirtland.af. mil/.

Nullmeyer, R.T., Stella, D., & Montijo, G. (2005). *Mishap trends in several Air Force aircraft: Implications for CRM traiing.* In Proceedings of the 11th International Symposium on Aviation Psychology. Oklahoma City, OK.

Rumsfeld, D. (2003) *Reducing Preventable Accidents.* Memorandum for Secretaries of the Military Departments. The Secretary of Defense, Washington DC.

Shappell, S.A. & Wiegmann, D.A. (2000). *The human factors analysis and classification system-HFACS.* (DOT/FAA/AM-00/7). Washington DC: Office of Aviation Medicine.

Wilson-Donnelly, K.A. and Shappell, S.A. (2004). *U.S. Navy/Marine Corps CRM training: separating theory from reality*. In Proceedings of the Human Factors and Ergonomics Society, New Orleans, LA.

COMPETENCE-BASED ASSESSMENT DESIGN FOR AIR TRAFFIC CONTROL TRAINING

Esther Oprins, **Ernst Burggraaff** & **Hans van Weerdenburg**
Air Traffic Control the Netherlands (LVNL)
Amsterdam, The Netherlands

Air Traffic Control (ATC) simulator and on-the-job training (OJT) requires a valid and reliable assessment system. Competence-based assessment results in more effective learning processes, better pass/fail decisions and improved selection criteria which may contribute to an increased output of competent controllers from training. This paper describes the design of the assessment system in use by Air Traffic Control the Netherlands (LVNL).

Introduction

Among other process control tasks in transportation and process industries, the ATC task is called a complex cognitive skill (Van Merrienboer, 1997). The processing of large amounts of dynamically changing information calls for complex cognitive processes. In combination with the strict safety requirements that do not allow any (human) error, this puts high demands on controller's competences. Selection requirements are generally very high, but often the outcome of training remains still too low, especially in ATC organizations serving busy and complex airports such as Schiphol Airport (LVNL). This may result in a shortage of controllers. Besides, high failure rates are undesirable because of the time-consuming and expensive simulator training and OJT. Solutions can be sought in improved selection, training, or both. An important part of training design is assessment. A valid and reliable assessment system may contribute to an increased output of competent controllers in several ways. First, assessment is a base for adequate feedback which supports trainee's learning processes. Second, training can be adapted to the trainee's needs which may increase learnability for more candidates. Third, more accurate judgments lead to better founded pass/fail decisions and help to reduce false positives and false negatives during later training phases. Fourth, higher reliability of training criteria makes it possible to obtain higher predictive validities for selection instruments.

Unfortunately, despite of its importance, scientific research on assessment in ATC training seems to have been very limited. The scarce literature is restricted to descriptions of ATC assessment from a more practical perspective (e.g. Hopkins, 1995). Some studies on ATC selection are relevant, because they involve assessment in work samples or criterion development for validation (e.g. Ramos, Heil & Manning, 2001). Within the field of aviation a substantial amount of research has been done on assessment of aircrew (e.g. O'Connor et al, 2002). The ATC task, however, is different. Its complex and time-critical character makes assessment extremely difficult to design. The invisible, cognitive processes result in subjective judgements of assessors, who mostly depend on over-the-shoulder observation.

The aim of this paper is to describe the design of the assessment system in ATC training at LVNL. We discuss how assessment based on competences may lead to more effective learning processes, better pass/fail decisions and improved selection criteria.

Competence-based Assessment: Background

The assessment system is based on principles of competence-based assessment found in the literature.

Competences

We define competence in relation to training as follows: 'the ability to apply acquired knowledge, skills and attitudes while performing tasks in realistic settings'. Competences are the result of a learning process for which a person needs specific abilities, personality and other features included in selection requirements (Roe, in press). Competences become implicit by learning since information processing has been automated. Although competences are not innate, they differ in trainability. Schneider (1990) separates 'consistent' components that improve by practicing from 'non-consistent' components that do not necessarily improve. This relates to 'recurrent' and 'non-recurrent' skills (Van Merriënboer, 1997).

Assessment of Competences

Assessment in ATC simulator training and OJT is usually defined as 'performance assessment' (Wickens, Mavor & McGee, 1997). An assessor judges performance criteria on a rating scale on the base of over-the-shoulder observation. These criteria are generally formulated in observable behavior, also called 'behavioral markers' (O'Connor et al, 2002). We consider competence-based assessment to be a specific type of performance assessment due to its focus on competences. In accordance with modern learning theories (Pellegrino, Chudowski & Glaser, 2001), competences are not analytically split up in detailed skills and knowledge, but assessment takes place at a higher level. This allows for different learning curves: (sub)skills and pieces of knowledge may be learned in a different order or tempo, as long as the competences required are obtained after a certain (flexible) learning period. Further, due to their more generic character, the same competences can be assessed during training. For instance, planning is relevant in each ATC training phase as well as in each ATC task execution. This makes it easier to identify trainee's strengths and weaknesses and to define appropriate training interventions at an early stage. In addition, progression on each competence can be measured, providing an important indicator of whether a trainee is still learning. This is essential in pass/fail decisions: when a trainee has not reached a learning plateau yet, it may be useful to continue training. In this respect, trainability of ATC competences must be taken into account (Schneider, 1990): consistent components (e.g. radiotelephony) that are not mastered yet could still be improved in contrast with non-consistent components (e.g. planning) that are more often reasons for failing.

In order to get a complete picture, covering the cognitive, emotional and social aspects, assessment should involve all these aspects that belong to a competence. In assessment of aircrew these are referred to as 'non-technical skills' (O' Connor et al, 2002). Many of them, such as situational awareness and decision making, are also essential for ATC.

A crucial step in the design process is a competence analysis. Involvement of air traffic controllers as subject matter experts (SME's) is extremely important because their implicit knowledge has to be explicated as the reverse process of learning.

Cognitive Processes

The assessment of cognitive processes is extremely important in ATC. This calls for an inference from observable behaviour and interaction with the trainee (e.g. asking questions). Feedback is more effective when coaches have insight into trainee's thinking patterns and strategies, which needs more emphasis in ATC training (Schneider, 1990). Besides, assessment of cognitive processes is required to obtain diagnostic information on performance shortcomings and to predict future performance in ATC training (Regian & Schneider, 1990).

The importance of assessment of cognitive processes is one reason why 'automated measurement' in ATC simulator training has hardly been applied, although safety and efficiency aspects such as separation, conflicts, delay and communication can be logged by the computer (Wickens, Mavor & McGee, 1997).

Selection

Competence-based assessment may indirectly contribute to better training results by using the competences as criteria in selection, because they optimally reflect the personal basis of job performance (Roe, in press). Performance measures obtained in training must be reliable and valid to make validation research valuable. Further, in work samples, as part of selection systems, competences can be rated using similar performance criteria as in training, serving as predictors for future performance. Thus, many similarities can be found between performance criteria applied in ATC training and in work samples (e.g. Ramos, Heil & Manning, 2001).

Psychometric Requirements

A precondition for any assessment system is its psychometric quality. Reliability and validity can be obtained by judgments of multiple assessors, assessor training, sophisticated measurement techniques (e.g. 'behavioral anchored scales'), representative tasks and performance criteria and so on (Berk, 1986).

Design Method

A competence analysis and literature research resulted in the ATC Performance Model which has served as a framework for the assessment design.

Competence Analysis

We organized two competence workshops in which twelve air traffic controllers formulated a set of thirteen competences. The set consists of: situational awareness, decisiveness, dealing with unexpected situations, workload management, conflict solving, multitasking, prioritizing, co-ordination and communication, flexible planning, leadership, teamwork ability, perseverance, and critical attitude. Each competence is supported by a set of eight to twelve behavioral markers. The collaboration tool Meetingworks was used, which makes it possible to brainstorm, discuss and structure electronically. This method enabled controllers to come to agreement about the completeness and the interpretation of each competence with aid of the behavioral markers, formulating them in their own jargon. This makes the competences recognizable and practically usable in training. Controllers were forced to think about their own work performance at a more abstract level.

Besides, we did literature research looking for additional aspects of ATC performance that might have been forgotten in the workshops. We were also interested in the relations between these aspects in order to categorize the thirteen competences. Thus, we compared our set with existing ATC (cognitive) task analyses (e.g. EATMP, 1999), performance models (e.g. Hadley, Guttman & Stringer, 1999), and performance measurement systems (e.g. Ramos, Heil & Manning, 2001). On the basis of this we developed the ATC Performance Model (Oprins & Schuver, 2003), presented in figure 1.

The ATC Performance Model

The model shows the dominant role of information processing in ATC work. Information processing provides the basis for actions, which result in the outcome: handling of air traffic. The way in which this happens depends on a number of influencing factors. All components of the model are specified in terms of competences. We recognize the majority of the competences defined in the workshops in the dark gray parts, some in the white parts. Some were revised and others added as a result of the additional literature research. We see that information processing comprises *situation assessment*, *planning* and *decision making*. This is mainly derived from the ATC model of Hadley, Guttman and Stringer (1999), but these cognitive processes are not necessarily ATC-specific in contrast with the actions and outcome. Situation assessment (e.g. Endsley, 1995) is further divided into *perception*, *attention management* and *interpretation* (mental picture). The actions consist of *communication*, *co-ordination*, *strip/label management*, and *equipment operation*. The outcome distinguishes *safety* and *efficiency*. The influencing factors are mainly represented by *workload management* and *teamwork ability*.

Properties of the Assessment System

Performance Criteria

Setting new performance criteria was the most fundamental change in the previous assessment system. They are directly derived from the ATC Performance Model. Each criterion is rated on a 6-points rating scale that strictly separates sufficient from insufficient behavior. A set of related criteria, formulated in terms of observable behavior, form a category representing a specific competence. Each category is visible in the model as a dark gray part and is marked in italics. The typical Dutch jargon proposed in the workshops has been maintained in order to maximize recognition and comprehension by the users. Most criteria are literally identical to the behavioral markers formulated in the workshops. The same fourteen categories are used for all ATC functions (e.g. area, aerodrome control), from the start of initial training till final job performance. They are even applied in two work samples that are part of the LVNL selection system. The criteria are also identical for each ATC function when possible, for instance, criteria of the category *communication*:

Communication *(all ATC functions)*
- Applies (non) standard phraseology correctly
- Express himself clearly, unambiguously and shortly
- Provides correct and sufficient traffic information

Only the criteria that belong to *safety* and *efficiency* are different because they have another meaning in each ATC function, illustrated by next example:

Safety *(ground control)*
- Prevents runway intrusions
- Arranges conflicts and right-of-way situations
- Checks correctness of clearances on strips and EDD
Safety *(area control)*
- Maintains separation minima correctly
- Builds in sufficient safety buffers
- Switches from monitoring to vectoring in time

The use of the same categories and performance criteria makes it possible to follow trainee's progression on each competence in order to define appropriate training interventions, based on trainee's weaknesses. The criteria can be applied in different task situations which provides a complete picture about trainee's performance. They are independent of variables such as traffic complexity or specific events which are relevant for the design of assessment tasks.

The ATC Performance Model provides indications on how the performance criteria can be assessed. First, the model separates objectively measurable criteria (outcome, actions) from criteria that can only be assessed subjectively (information processing). We have argued that safety and efficiency could even be 'automatically' measured in the simulator. This distinction is useful for assessors who have to be aware of their own restrictions when they depend on subjective measures. Second, the model gives information about trainability. Actions are trainable because they improve by practicing in contrast with the majority of the cognitive processes. The latter express the 'gut feeling' of assessors. They help them to argue why trainees perform (in)sufficiently as causes for (in)sufficient actions or outcome, relevant for adequate feedback and pass/fail decisions.

Phasing

We divided the training period into phases and determined performance standards to be achieved at the end of each phase. Trainee's competence is assessed against lower standards in earlier phases. Before the introduction of phasing assessors did only rely on their experience, which increased disagreement between them. Trainees did not have insight into the standards required in final or in intermediate phases. This vagueness did certainly not contribute to learning and to succeed in training.

In simulator training, phases are mainly defined by the sequence of simulator scenarios (Farmer et al, 1999). OJT normally occurs 'unstructured', not only in ATC (Jacobs & Jones, 1995). Structuring OJT is difficult because assessment tasks cannot be arranged due to the ongoing live traffic. We divided OJT in phases based on three basic principles: degree of *safety/efficiency*, *complexity* of traffic situations, and *independence* of the coach. Each OJT consists of four phases with flexible lengths, dependent on trainee's assessment results (progression), to take into account individual differences in learning. An example of OJT phases in area control is the following:

Phase 2 (8-14 weeks): to be able to handle *standard* traffic both *safely* and *efficiently*, and *complex* traffic *safely*, independently of the coach.
Phase 3 (8-14 weeks): to be able to handle both *standard* and *complex* traffic both *safely* and *efficiently* independently of the coach

Standard and complex traffic are further detailed in terms of traffic intensity, diversity in aircraft, flight destinations, weather circumstances, runways in use, specific events and so on. These variables are predetermined in simulator scenarios (assessment tasks), but in OJT only a description for trainees and assessors is available serving as a guideline. Safety and efficiency, as well as the other competences, are specified for each separate ATC function and for each phase in both simulator training and OJT. Therefore the performance criteria are accompanied by behavioral examples, illustrated in figure 2, which can be considered as a variant on 'behaviorally anchored scales' (Berk, 1986). The examples do not specify the scale positions, but represent the performance standards to be achieved at the end of each phase. Differences between ATC functions, also for the criteria that stay the same (e.g. planning, communication), become directly visible in these examples, which are necessarily function-specific to be as clear as possible. The examples contribute to consistent judgments between assessors, not only for assessing against the same performance standards in a phase but also for assigning specific behavior to the same criterion. For trainees it is clearer what is expected from them in a specific training phase.

Continuous Assessment

Continuous assessment is applied as in the majority of ATC organizations (Hopkins, 1995). Coaches, who are also assessor, measure trainee's achievement during training. They are continuously in interaction with the trainee and can force trainees to verbalize their thoughts. This enables them to assess cognitive processes (e.g. strategies). Representativeness of task situations is guaranteed, because assessment is not restricted to a particular moment. Multiple assessors are involved for maximizing reliability, who are trained beforehand in the use of the system, interpretation of performance criteria, and avoidance of rating errors. However, objective measurement is impossible since coaches are constantly influencing trainee's task performance by their guidance. Therefore, we combine continuous assessment with performance tests (Berk, 1986).

Performance Tests

Performance tests measure trainee's performance during a test in the simulator or in OJT without coaching interventions. We emphasize the objective character of the test as a counterpart of continuous assessment in several ways. In the simulator checklists are used for the observation of events occurring in scenarios at a specific time (e.g. conflicts, runway changes), added by possible solutions for each event. The solution chosen by the trainee is marked. Afterwards the final test score is calculated. This final score is the sum of weighted scores that are assigned to each criterion. The weighting relates to the ATC Performance Model: the sum of the scores belonging to information processing has the same weight as the actions and outcome together, because information processing refers to the causes for (in)sufficient actions and outcome. Safety is measured objectively by counting the number of safety violations (e.g. unsolved conflicts), based on the annotations on the checklist.

Conclusions and Future Directions

The assessment system has been used in practice for two years now. The evaluation of the system comprises several parts. First, we investigated the practical use of the system and the improvement of learning processes for coaches and trainees qualitatively (interviews, questionnaire, report analysis). This has led to positive results. The involvement of controllers in the design process has contributed to a better recognition of behavior. Assessors more easily express and validate their 'gut feeling' with aid of the competences, which results in more appropriate training interventions and better founded pass/fail decisions. The ATC Performance Model helps them to get insight into the different components of performance, such as the distinction between objective and subjective measures and the extent of trainability of competences. The use of the same performance criteria makes assessors more familiarized with their meaning. It also helps them to follow trainee's progression on each competence and to provide adequate feedback. Agreement about performance standards in different phases is higher. For trainees it is clearer which competences they must develop further in a specific phase. Thus, the assessment system is definitely practically usable and contributes to more effective learning processes.

Second, we are investigating the psychometric quality of the system, especially the interrater reliability, internal consistency, and predictive validity. However, more long-term evaluation is needed for quantitative conclusions about a possible increased output from training, although the findings about improved learning processes are encouraging. This evaluation research has to be regarded in relation to the selection system, which we have redesigned simultaneously using the competences as criteria, and other possible influences (e.g. changes in training design). This makes it all rather complex. For facilitation of this further research we make use of a database that stores all selection and training results. Therefore, assessors fill in assessments digitally by means of the web-based assessment tool Questionmark Perception. This tool has several advantages, not only for research purposes. First, trainee's progress can be better followed by interested persons (e.g. training managers) who have access from several places so that interventions can be undertaken as soon as possible. For instance, from the office there will be direct access to assessments that takes place at the tower. Second, different overviews and graphs (e.g. individual learning curves) can be distillated from the system, which provides more insight in learning processes. Third, reliability is increased because the system forces assessors to fill in assessment reports accurately and univocally. Finally, training results can more easily be used for validation studies for both selection and training purposes, needed for long-term evaluation. This research will be an on-going process which makes it possible to adapt performance standards in training and selection requirements continuously in order to maximize output from training ultimately.

References

Berk, R.A. (1986). *Performance assessment: methods and applications.* The Johns Hopkins University: Baltimore.

EATMP (1999). *Integrated task and job analysis of air traffic controllers phase 2: task analysis of en-route controllers* (HUM.ET1.ST01.1000-REP-04). Luxemburg: Eurocontrol.

Endsley, M. (1995). Towards a theory of situational awareness in dynamic systems. *Human factors, 37*, 32-64.

Farmer, E., Rooij, J. van, Riemersma, J., Jorna, P., & Moraal, J. (1999). *Handbook of simulator-based training.* Aldershot: Ashgate.

Hadley, G., Guttman, J., & Stringer, P. (1999). *Air traffic control specialist performance measurement database* (DOT/FAA/CT-TN99/17). Washington DC: FAA.

Hopkins, D. (1995). *Human factors in air traffic control.* London: Taylor & Francis.

Jacobs, R.L., & Jones, M.J. (1995). *Structured on-the-job training. Unleashing employee expertise in the workplace.* San Francisco: Berrett-Koehler Publishers.

Merrienboer, J.J.G. van (1997). *Training complex skills: a four component instructional design model for technical training.* Englewood Cliffs, NJ: Educational Technology Publications.

O'Connor, P., Hormann, H., Flin, R., Lodge, M., & Goeters, K. (2002). Developing a method for evaluating crew resource management skills: a European perspective. *The international journal of aviation psychology, 12,* 263-285.

Oprins, E., & Schuver, M. (2003). Competentiegericht opleiden en beoordelen bij LVNL. *HUFAG Nieuwsbrief, 6,* 2-4.

Pellegrino, J.M., Chudowski, N., & Glaser, R. (2001). *Knowing what students know. The science and design of educational assessment.* Washington: National Academic Press.

Ramos, R.A., Heil, M.C., & Manning, C.A. (2001). *Documentation of validity for the AT-SAT Computerized Test Battery: volume II* (DOT/FAA/AM-01/6). Washington, DC: FAA.

Regian, J.W., & Schneider, W. (1990). Assessment procedures for predicting and optimizing skill acquisition after extensive practice. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.). *Diagnostic monitoring of skill and knowledge acquisition* (pp. 297-323). Hillsdale, NJ: Lawrence Erlbaum Associates.

Roe, R.A. (in press). The design of selection systems: contexts, principles, issues. In A. Evers, O. Smit, & N. Anderson (Eds.), *Handbook of personnel selection*. Oxford: Blackwell.

Schneider, W. (1990). Training high-performance skills: fallacies and guidelines. In M. Venturino (Ed.), *Selected readings in human factors* (pp. 297-311). Santa Monica, CA: The human factors society.

Wickens, C.D., Mavor, A.S., & McGee, J.P. (1997). *Flight to the future: human factors in air traffic control*. Washington: National Academy Press.

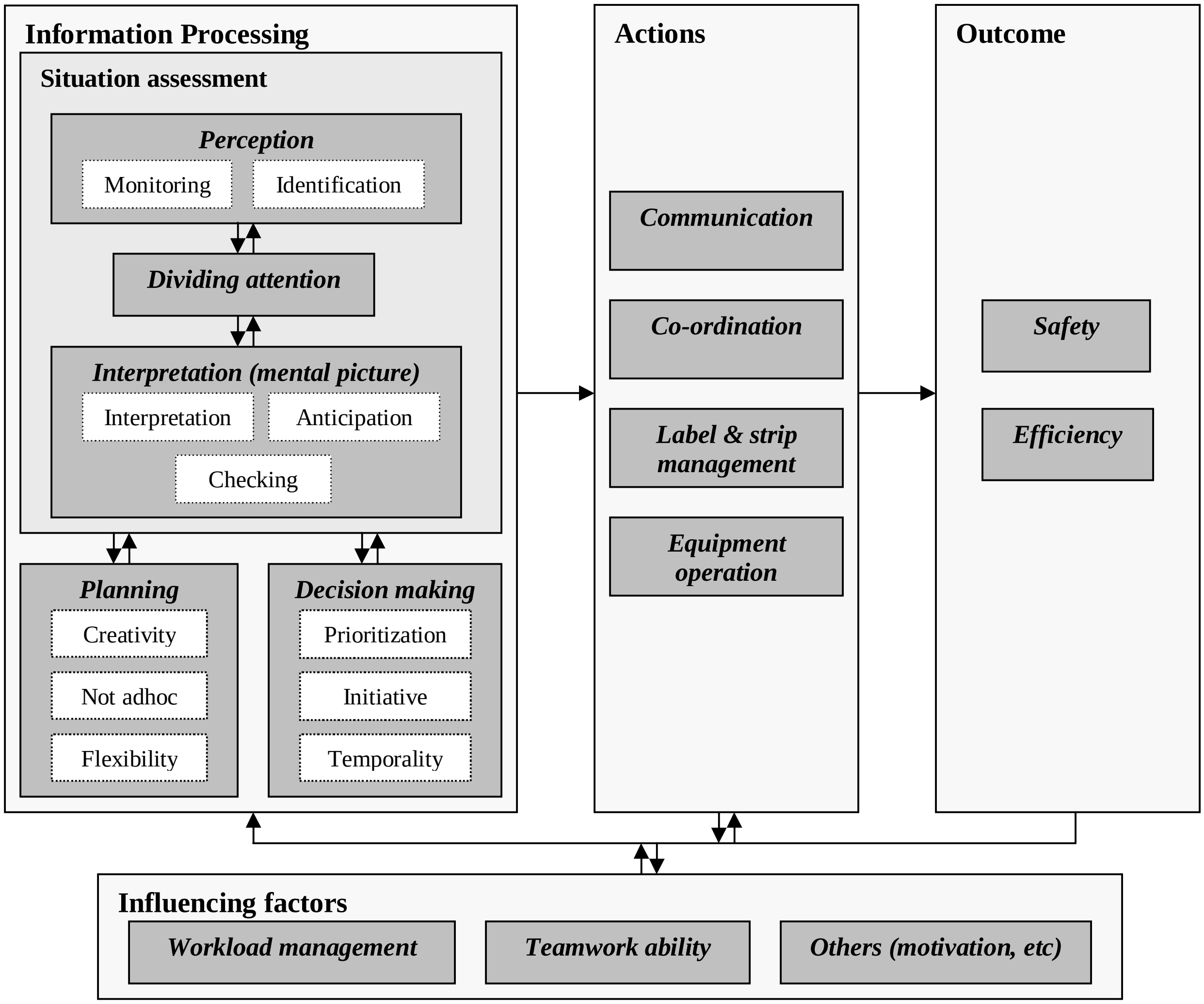


Figure 1. *The ATC Performance Model (Oprins & Schuver, 2003)*

Efficiency	**Phase 1**	**Phase 2**
• Applies speed control correctly • Applies vector technique correctly • Takes into account aircraft performances • Takes into account different flight levels • Builds a sequence of climbing and descending traffic	Speed control and vector technique does not need to be optimal, but must be conflict free and conform standard routes and transfer, taking into account differences in aircraft performances and in time turning to own or published navigation or speed. Some delay may still occur; sequences are not always efficiently enough yet. Application of level separation and assessment of intermediate levels during sequencing is not always optimal, for instance, 2 flights in different STARS are cleared to the same FL.	Sequences consist of 5 to 7 NM interval, with minimal speed differences by optimal speed control of inbounds, and with parallel handling over of outbounds by optimal vectoring. There is a striving for continuous climb/descend, taking into account differences in aircraft performances and in time turning to own or published navigation or speed. Delay has been avoided whenever is possible. Level separation and assessment of intermediate flight levels during sequencing is applied optimally.

Figure 2. *Performance standards in two phases of area control training for Efficiency.*

Integrated Automation Enhances Air Traffic Controller Conflict Detection Performance Under Free Flight

Raja Parasuraman
George Mason University, Fairfax, VA, USA

Ulla Metzger
Deutsche Bahn (DB) Systemtechnik, Munich, Germany

Previous simulation studies have shown that, without the assistance of automation, controllers have difficulty in timely detection and resolution of aircraft-to-aircraft conflicts under future air traffic management concepts such as Free Flight (Galster et al., 2001; Metzger & Parasuraman, in press). However, how automated information on potential conflicts should be displayed to the controller is not well understood. In the present study, we reasoned that automation that was integrated with the primary radar display would be effective in enhancing conflict detetction. We also manipulated display modality so that the automation provided information to the controller using simple visual, enhanced visual, or visual+auditory (multi-modal) displays. It was expected that under high traffic load, the associated requirement for communication by datalink could divert visual attention from the radar, thus potentially impairing conflict detection performance and necessitating automated assistance. We hypothesized that multi-modal feedback could lead to earlier conflict detection than purely visual feedback by better guiding visual attention. A performance benefit was also expected for the enhanced visual aid by providing more transparency regarding conflict prediction and reducing the requirement for visual search.

Eight experienced, full-performance level controllers were tested on an ATC simulator displaying a generic airspace and consisting of a radar display, a datalink display, and electronic flight strips, presented on two different 21-inch monitors. Traffic density was manipulated to be either moderate (on average about 10 aircraft in a 50-mile radius sector) or high (about 16 aircraft). In addition, the feedback type of the conflict detection aid was varied. In the simple condition, two red filled circles predicted which aircraft pair would be in conflict. In the enhanced visual condition, the circles were supplemented with two red heading lines indicating why and where the aircraft were predicted to be in conflict. Finally, in the multi-modal condition, the enhanced visual aid was supplemented by an auditory alert presented on loudspeakers on either side of the monitors. In addition to performance and subjective measures, ocular activity (i.e. fixations and dwells) was recorded with an ASL Model 5000 head-mounted eye tracker at a sampling rate of 60 Hz as a measure of visual attention.

Of the several results of interest, a few are reported here. First, the present experiment provided additional evidence that controller performance under Free Flight can be improved with the help of effective automated decision aids. Conflict detection performance was substantially improved—to near perfect performance—by the automated aids. However, the prediction that multi-modal feedback would result in better conflict detection performance (especially earlier detection) than simple visual feedback was not supported. The expected differential benefit of the visually enhanced feedback was also not found. Controllers fixated over 60% of the time on the radar display, which may explain why no differential effects of the automated aids were found: when attention is allocated to the radar most of the time it is unlikely that a salient visual aid is missed, and enhancing the visual aid or adding redundant auditory information provides no additional benefit. Finally, the sizeable benefits provided by the automated aids may largely due to the automation being integrated into the primary radar display, which was the major focus of controller attention.

THE STL MODEL: A THREE-DIMENSIONAL PERSPECTIVE ON ORGANIZATIONAL CHANGE PROGRAMS

Manoj S. Patankar, Ph.D.
Saint Louis University
St. Louis, Missouri

Thomas Bigda-Peyton, Ph.D.
Action Learning Systems
Boston, Massachusetts

Jeffery Brown, M. Ed.
System Safety Group
Peterborough, New Hampshire

This paper examines organizational change programs across aviation, healthcare, and financial services sectors. Based on the analysis of three key programs, a theoretical model, which could be used to describe the state of any organizational change program, is presented. This model is called the STL Model and is represented by three mutually perpendicular axes: scalability, transferability, and longevity. In simplest terms, scalability refers to "volume;" greater the volume of users of, or participants in, a particular change program, the greater the scalability of that change program. Transferability refers to the number of user-groups, whether within a specific discipline or outside. Longevity is the measure of how long a particular change program is in operation or existence. Although longevity alone does not necessitate progress along the scalability and transferability axes, it does provide an opportunity for improved scalability and transferability. It is hypothesized that certain factors, called "influence vectors" could be managed to improve the overall sustainability of organizational change programs.

Background

Literature on organizational learning underscores the importance of systems thinking in designing and managing change efforts (cf. Senge, 1990; Senge, Kleiner, Roberts, Ross, and Smith, 1994; Senge, Kleiner, Roberts, Ross, Roth, & Smith, 1999). However, studies addressing the difficulties in scaling "n of 1" type innovations within an organization or across multiple organizations have been limited.

Three examples, one from aviation, one from health care, and one from financial services, are presented in the following section. These examples illustrate specific "n of 1" innovations that the research team has reported in previous research. Such innovations or success stories abound in several industry segments, but most of them never scale high enough for organization-wide adoption. In this paper, a theoretical model—called the STL Model—is presented as a means to characterize the progress of change programs in terms of scalability, transferability, and longevity and also to present the role of "influence vectors" in transforming a change program into a lasting, institutionalized cultural change.

Examples from Three Distinct Industries

Example 1: Application of the Concept Alignment Process in aviation maintenance

The Concept Alignment Process (CAP) was first adopted by the subject aviation department's flight crew in 1995 and subsequently customized by the maintenance department to suit their needs. CAP is different from most of the Maintenance Resource Management (MRM) programs because it focuses on a behavioral change rather than an attitudinal change. This program illustrates that an organization need not change everyone's safety attitude before expecting a change in anyone's behavior. CAP expects all employees to change their behavior and follow a prearranged process. Therefore, it does not suffer from the limitations of the first three generations of MRM programs: limited success in achieving behavioral changes, changes lasted for six to twelve months after the training, and participants' attitudes toward the program declined over time (cf. Taylor & Christensen, 1998; Taylor & Patankar, 2001).

Patankar and Taylor (1999) reported that not all of the technicians in this organization practiced CAP to the same extent. For example, some of them understood the basic protocol, but hesitated to challenge another person's concept or to seek validation. Only a few individuals were observed to be practicing CAP consistently and to its full potential (challenging concepts, seeking validation, identifying causes for ambiguity in information, and implementing appropriate structural/procedural changes so that the ambiguities are minimized). Patankar and Taylor also observed that as the skeptics used the process, they understood it more clearly; and as their success in effecting organizational changes grew, their trust in the process grew. Gradually, they were becoming believers. Hence, this company was able to cause an attitudinal change through a behavioral change, rather than the other way around (as attempted in the previous three generations of MRM programs).

Example 2: Collaborative Rounds—An Interdisciplinary Innovation in the Post-surgical Care of Open-heart Surgery Patients.

Caring for open-heart surgery patients is a socially and technically complex endeavor. Surgeons, therapists, nurses, pharmacists, social workers, and many other disciplines must coordinate their assessments and actions with one another, and with patients and their families. Depending on treatment needs, up to fifteen different disciplines may independently gather information from the patient on any given day (much of it is redundant), develop a care plan, and enter the plan into the patient's medical chart. The medical chart is typically the primary means of coordinating the thoughts and actions of all care providers working with the patient. Yet, the written record is an imperfect means of coordinating activity; informational gaps, ambiguous data entry, changes in patient status and other issues routinely compel care providers to seek or provide clarifying information in order to fit together the patient's total care plan. This approach is inefficient and prone to oversights and conflicting actions based on misunderstandings of the patient's situation and uncertainty about the overall plan of care.

In 1999, clinicians concerned with these problems at an acute care hospital began re-thinking the post-surgical care processes for open-heart surgery patients. These care providers, eventually known as the cardiac surgery team, decided that altering the patterns of interaction and communication would be integral to improving the processes of care for their patients. They decided to collaborate, to bring all disciplines together at the same time each morning, and to partner with patients and their families in assessment and care planning (cf. Uhlig, Nason, Camelio, Kendall, & Brown, 2002).

Adapting team briefing and debriefing strategies from the air transportation industry, the team achieved significant reduction in operative morbidity and mortality, as well as substantial gains in staff and patient satisfaction. An important feature of the Concord Collaborative Rounds model was the deliberate capture, through debriefing, of "systemic glitches"—deviations from intention (errors) that could support identification, analysis and intervention in hazards and error-provoking conditions. Despite significant reduction in operative morbidity and mortality and national recognition of this change in practice as an important safety innovation, this innovation was not sustained by the organization following a change in physician leadership.

Example 3: A Change Program in Financial Services

In 1999, the COO and CIO of the fixed-income division of a major investment firm recognized that they were reaching the limits of their current model of operations. Along with the CEO and other senior executives, they had started a long-term organizational change process in the early 1990's. When they began, the division had just been embarrassed by a major error (and resulting loss of business) that occurred when an investment professional took "too much risk." A new CEO was brought in, and he decided to take significant steps to "clean up the mess." The "mess" involved radical decentralization of the professionals, to the point where (reportedly) the division had "120 professionals and 120 Information Technology (IT) systems." They also had that many approaches to investment decision-making.

The new CEO began by recruiting two key managers from the professional side to create more coherence across the unit. They created an approach called TAM ("Targeted Active Decision-Making") that placed boundaries, or limits, on the range of acceptable risks for investment decision-making. This template was vigorously enforced and reduced variations in professional practice. Next, the CEO chartered (and the COO led) a redesign of all workflows related to investment decision-making, using a reengineering approach. This led to improved productivity and efficiency, on the one hand, and to strengthening the lines of business (or "desks") on the other. By 1998, the CEO was promoting the third phase of the change—reconciling the IT systems, which had remained untouched throughout the preceding phases. The IT initiative was based on rationalizing the infrastructure; a key mechanism was the concept, and principle, of "reusable components." The concept was simple but hard to implement. IT project teams, as well as the overall IT management team, were tasked with transferring and re-using "good" components from one team, or line of business, to another. This was hard for several reasons: it was unfamiliar; it represented a significant cultural change for the IT managers, their teams, and the lines of business they supported; and it meant overcoming the idea that every line of business is unique—that is, one cannot (by definition) re-use a component built for the Bonds desk in a software tool for the Money Market desk.

The change effort began with an action science-based (cf. Argyris, Putman, & Smith, 1985; Argyris, 1992) approach called action learning (or "Active Learning" as it was called by the team), using After Action Reviews of situations and stories from projects. The research team, led by Dr. Bigda-Peyton, used an appreciative inquiry orientation, in which they began with successes, mapped out the actions and operating assumptions that led to the successes, and then (and only then) looked at "unintended consequence." (Even then, they began with unintended positive consequences). From this starting point, they constructed an "As Is" and a "To Be" that were, in effect, a picture of the current work culture and a more desirable (and effective) work culture that they could create together, and with their business counterparts. For instance, they agreed that a key driver of their current culture was the "hero model"—relying on individual experts to solve critical problems (Bigda-Peyton & Galor, 1999). They further agreed that they wanted to create a culture driven by "shared accountability with individual excellence." This and other, similar drivers became the metrics by which they evaluated the change program.

The second phase of the intervention used peer reviews, in-action problem-solving, and surfacing and using tacit knowledge of the business landscape and the software development process. The research team made a breakthrough on the re-usable component issue; the team got a major win and gained significant credibility with the business side as a result. The lead technology architect commented, "I didn't know you could solve a technology problem with a model like this—we didn't learn this in software engineering school!" The team also used the method to solve other key problems, such as the departure of the lead architect and how to facilitate a project to solve the problems of a desk that were notorious for "broken" processes and uneven results. The team used the approaches to make significant strides in all of these areas.

In the third phase of the effort, the team engaged the business side as well as their immediate allies and partners in the central IT organization. They also did a parallel project with the central Risk and Knowledge Management group. In each case, the work was well received; but after one "handshake" project, the effort declined. The initiative subsided in 2002, after measurable and significant gains in innovation, operations effectiveness, customer satisfaction, and culture shift. The specific reasons for the erosion of this change program were never formally investigated. (cf. Bigda-Peyton, 2002).

These examples illustrate the following: (a) change programs are heavily influenced by their specific champions and (b) the resistance to change among field personnel could go as far as a high degree of professional rejection of persons championing the change. However, it is not clear what specific strategies change agents need to employ in the design and implementation of their programs in order to increase the probability of institutionalization of those programs. Additionally, the specific "incubation time" should be identified for change programs before their effects are evaluated. Although similarities and differences in terms of professional, national, and organizational cultures among physicians, pilots, and aircraft mechanics have been reported in past studies (Helmreich & Merritt, 1998; Taylor & Patankar, 1999), the specific roles of such cultures in influencing the propagation of an organizational change program have not been examined.

The STL Model: Characterization and Analysis of Organizational Change Programs

Although only three specific cases are described above, the underlying problems of scalability, transferability, and longevity seem to be consistent in many similar cases. These problems are robust—they seem to exist across sectors and methods of intervention. Thus, the researchers believe that there must be some fundamental issues that need to be addressed. Therefore, it appears that the problems of scalability (S), transferability (T), and longevity (L) could be framed in the form of a three-dimensional model—called the STL Model (see Figure 1)—that could then be used to assess the success of a previous intervention (retrospective analysis) or to develop specific strategies to ensure sustainability of future or planned change programs (prospective analysis).

The STL Model could allow—for the first time—researchers, practitioners, and policy makers to view change programs in terms of three interrelated dimensions. By bringing scalability, transferability, and longevity perspectives together, one could begin to formulate a fresh and integrated view of the assessment of change. Further, it is postulated that this view includes new micro-level (individual or small-group) dimensions that could promote the understanding of the dynamics of knowledge transfer/flow during the progression of a change program and the effect such flows, as well as people or "nodes in a network" that are responsible for knowledge transfer, might have on the overall success of the change program. Finally, the STL Model could enable tracking of the factors contributing to the dissemination (or lack thereof) of local innovations on a wider scale. In turn, an enhanced perspective on the assessment of organizational change could enable the promotion of the spread of innovation and, in part, help address the fundamental problem of transfer of innovation.

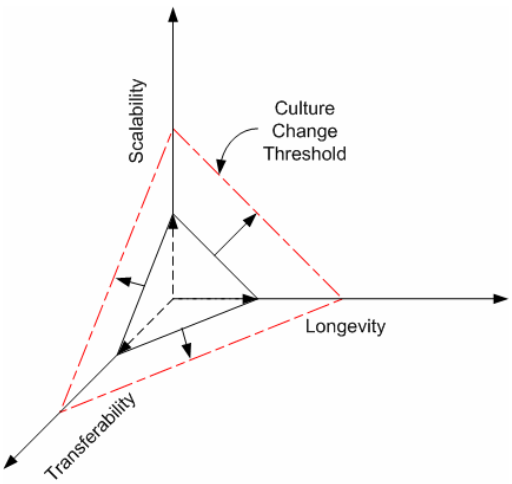


Figure 1. *The STL Model*

The Three Axes

Scalability, transferability, and longevity are three inter-related dimensions that could be expressed along three mutually perpendicular axes. Clearly, a change in one of the dimensions could affect the other two. However, it is important to note that a certain degree of progress along all three dimensions is necessary for a change program to achieve the desired level of sustainability. Therefore, one could argue that scalability, transferability, and longevity collectively define the sustainability of a change program.

In simplest terms, scalability refers to "volume." The greater the volume of users of, or participants in, a particular change program, the greater the scalability of that change program. For measurement, researchers could count the number of users and estimate the program's scalability.

Transferability refers to the number of different user-groups, whether within or between disciplines. For example, the transferability of a pre- and post-task briefing process could be measured in terms of the number of work groups using the process. As the number of work groups increases, the transferability increases. If the transferability goes beyond the traditional boundaries of an organizational unit, the change could be expressed in terms of orders of magnitude—when the briefing process that was first used in the flight operations department is adopted by the maintenance department, there is a jump in transferability by one order of magnitude. If the same process is adopted by another department, the transferability of that process will undergo a jump by another order of magnitude. A multi-order transferability is possible when that process is adopted by an entirely different industry such as healthcare. Obviously, as the transferability increases, so does the scalability; however, change along this dimension is most difficult.

Longevity is simply the measure of how long a particular change program has been in existence. Longevity alone does not necessitate progress along scalability and transferability because organizational silos could keep a certain change program alive and hidden for a long time. Nonetheless, longevity does provide an opportunity for improved scalability and transferability. From another perspective, a certain degree of longevity is essential for an innovation to be visible outside of a particular organizational unit. Furthermore, for a bi-directional transfer to take place, the innovation has to last in the originating organization long enough for the new organization to adopt it, demonstrate the advantages, and report unique, applicable findings back to the original organization.

The Influence Vectors

The influence vectors are specific individual or organizational factors such as participant attitudes, management tenure, regulatory requirements, labor-management relationship, etc. that have a significant influence on the lifecycle of a particular change program. Such factors are called "vectors" because they have a magnitude and a direction: magnitude is quantified via opinion or attitude scales, or by quantitative archival evidence; direction is defined by the effect of that factor on the change program—if the effect is beneficial, the direction is positive. Also, it must be noted that a particular influence vector is likely to impact all three dimensions, and its influence may vary. For example, the tenure of a manager, measured in terms of years, may have a positive influence on the change program and thereby boost its longevity. The positive effects of such longevity (linear scale) might actually cause exponential changes on the transferability and scalability dimensions. Survey questionnaires and archival data analysis could be used to determine the key influence vectors in a particular industry segment. It is then plausible that influence vectors could be managed to drive specific change programs to their intended level of institutionalization.

The Culture Change Threshold

Differences in national, organizational, and professional cultures in aviation and health care have been reported (Helmreich & Merritt, 1998; Taylor & Patankar, 1999). Similarly, the role of organizational cultures in safety-critical industries has also been studied extensively (Reason, 1997; Westrum & Adamski, 1999). Largely, these studies have focused on describing the concept of culture or safety culture. Concurrently, many organizations have elected to implement system-wide changes; however, because the current state of knowledge mainly addresses the definition of culture, it is not clear when such organizational changes should be declared successful or when exactly one could declare that there has been a cultural change. By measuring a particular change program along three interrelated dimensions, the STL Model attempts to express the state of the change accomplished by the specific change program. Based on the literature that describes failures of various change programs, one could postulate that cultural change, as opposed to climatic change, is a long-term change in behaviors as well as attitudes of the individuals and it changes organizational structures, processes, and policies. Such a change eventually becomes independent of the initiating champion. Therefore, in all measures, a cultural change is not likely to relapse. If the state of a particular change program could be described in terms of scalability (the number of individuals using it), transferability (the number of organizational units using it), and longevity (the total years that it has been in existence), one may be able to define a three-dimensional threshold beyond which the change could be considered long-term enough to be commonly accepted as a cultural change.

Conclusion and Future Directions

In order to thoroughly test the STL Model, both retrospective as well as prospective analyses are essential. In the retrospective mode, the characteristics of previously implemented change programs—their scalability, transferability, and longevity need to be quantified. Also, it would be imperative to study the influence vectors as well as knowledge transfer nodes, both positive as well as negative, that affected the final status of the change program. It is important to conduct prospective analysis to determine what factors the industry partners believe would make a significant difference in the transfer of future innovations. Also, the prospective analysis allows for a critical window of opportunity to facilitate the transfer of innovations across organizational units or industry sectors.

The following hypotheses need to be tested:

- Hypothesis # 1: The state of a change program can be defined in terms of the three macro-level dimensions of the STL Model: scalability, transferability, and longevity.
  - Secondary Hypothesis: For each dimension of the STL Model, there are micro-level influence vectors that have either a positive or a negative effect on the development of the corresponding dimension.
- Hypothesis # 2: Transferability of innovation can be engineered across organizational units or disciplinary boundaries through appropriate control of the influence vectors.
  - Secondary Hypothesis: Transfer of innovation is influenced by nodes in organizational networks and the presence of a learning culture.

Both hypotheses, and their associated secondary hypotheses, could be tested in a cycle of data collection, analysis, and testing as the research progresses through three possible phases—single case in each sector, three-to-five cases in each sector, and seven or more cases in each sector.

Kramer and Sabin (2003) describe three conceptual phases of organizational learning (generating new knowledge, creating organizational memory, and embedding the learning) and present practical activities that professionals can use to promote learning to change organizations and influence key outcomes. Organizational learning techniques such as the After Action Review (AAR) could be employed to identify lessons learned from critical experiences (cf. Baird, Holland, & Deacon, 1999). Learning impediments described by research participants need be analyzed using a model developed by Shaw and Perkins (1992) that categorizes learning barriers in terms of insufficient capacities to reflect on experiences, disseminate knowledge, and/or take appropriate action. The Dimensions of the Learning Organization Questionnaire (DLOQ) by Marsick and Watkins (2003) could be used to assess the cultures of participating organizations. Empirical evidence demonstrates that the DLOQ measures seven dimensions (continuous learning, inquiry and dialog, team learning, empowerment, embedded system, system connection, strategic leadership) that impact learning, sustain change, and drive improved performance (Yang, Watkins, & Marsick, 2004).

In summary, the STL Model seems to offer a plausible means to characterize organizational change programs. Empirical research in multiple industries could be used to test the validity of this model.

References

Argyris, C., Putnam, R., & Smith, D. (1985). *Action science*, San Francisco: Jossey-Bass.

Argyris, C. (1992). *On organizational learning.* Malden, MA: Blackwell Publishers.

Baird, L., Holland, P., & Deacon, S. (1999). Learning from action: Imbedding more learning into the performance fast enough to make a difference. *Organizational Dynamics,* Spring, 19-32.

Bigda-Peyton (2002). *Active learning and risk management in software development*. Internal presentation and analysis, Fixed Income Division and Corporate Consulting Group, A corporate client.

Bigda-Peyton, T., & Galor, J. (1999). *Creating a New Culture: Heroes at Otis*. An internal presentation to the Global Product Strategy Design Team, Otis Elevator Company.

Helmreich, R., & Merritt, A. (1998). *Culture at work in aviation and medicine: National, organizational, and professional influences.* Aldershot, U.K.: Ashgate Publishing Limited.

Kramer, T.J., & Sabin, E.J. (2003). Managing the organizational learning process to improve the bottom line. *The Psychologist-Manager Journal*, 6, 11-30.

Leonard, D., & Swap, W. (2002). In Hesselbein, F., Goldsmith, M., Somerville, I. *Leading for innovation and organizing for results* (165-176). San Francisco: Jossey-Bass.

Marsick, V. J., & Watkins, K. E. (2003). Demonstrating the value of an organization's learning culture: The dimensions of the Learning Organization Questionnaire. *Advances in Developing Human Resources*, 5, 132-151.

Patankar, M., & Taylor, J. (1999). *Corporate aviation on the leading edge: systemic implementation of macro-human factors in aviation maintenance*. Paper published by the SAE, Aerospace Group [SAE Technical Paper Number 1999-01-1596]. Warrendale, PA.

Reason, J. (1997). *Managing the risk of organizational accidents.* Aldershot, U.K.: Ashgate Publishing Limited.

Rogers, E. (1983). *Diffusion of innovations.* New York: Free Press.

Senge, P. (1990). *The fifth discipline: The art and practice of the learning organization.* New York: Doubleday.

Senge, P.M., Kleiner, A., Roberts, C., Ross, R.B., & Smith, B.J. (1994). *The fifth discipline fieldbook: Strategies and tools for building a learning organization.* New York: Currency Doubleday.

Senge, P.M., Kleiner, A., Roberts, C., Ross, R.B., Roth, G., & Smith, B.J. (1999). *The dance of change: The challenge to sustaining momentum in learning organizations.* New York: Currency Doubleday.

Shaw, R., & Perkins, D. (1992). Teaching organizations to learn: The power of productive failures. In Nadler, D.A., Gerstain, M.S., & Shaw, R.B. (Eds.), *Organizational Architecture* (pp. 175-192). San Francisco: Jossey Bass.

Taylor, J.C., & Christensen, T.D. (1998). *Aviation Maintenance Resource Management: Improving communication.* Warrendale, PA: SAE Press.

Taylor, J., & Patankar, M (1999). Cultural Factors contributing to the successful implementation of macro human factors principles in aviation maintenance. In R. Jensen (Ed.) *Proceedings of the Tenth International Symposium on Aviation Psychology,* [CD-ROM], at Ohio State University, Columbus, OH.

Taylor, J., & Patankar, M. (2001). Four generations of MRM: Evolution of human error management programs in the United States. *Journal of Air Transportation World Wide 6*(2), 3-32. Omaha, NE: University of Nebraska at Omaha.

Uhlig, P., Nason, A., Camelio, A., Kendall, E., Brown, J. (2002). System innovation: Concord Hospital. Section on John M. Eisenberg Patient Safety Awards. *Journal on Quality Improvement,* Volume 28 (12).

Westrum, R., & Adamski, A. (1999). Organizational factors associated with safety and mission success in aviation environments. In D. Garland, J. Wise, & V. Hopkin (Eds.), *Handbook of aviation human factors* (pp. 67-104). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Yang, B., Watkins, K., & Marsick, V. (2004). The construct of the learning organization: Dimensions, measurement, and validation. *Human Resource Development Quarterly,* 15, 31-55.

TEAM COORDINATION IN UAV OPERATIONS

Harry K. Pedersen
Cognitive Engineering Research Institute
Mesa, Arizona

Nancy J. Cooke
Arizona State University East and
Cognitive Engineering Research Institute
Mesa, Arizona

The term "unmanned" in the context of unmanned aerial vehicle (UAV) operations is too often taken literally, overlooking the humans controlling, monitoring, collaborating, and coordinating from the ground. Promoting and improving the performance of the human component in the operation of UAVs is paramount and enhancing the coordination of the humans in the system is one of many important human factors issues which must be overcome. Research from the Cognitive Engineering on Team Tasks Laboratory has approached this problem with the development of a synthetic test-bed replicating UAV coordination in the lab. Findings from this synthetic task environment (STE) will be discussed in context of the implications that UAVs are in fact manned and require the attention of the human factors community.

Introduction

The Department of Defense defines unmanned aerial vehicles (UAVs) as powered aerial vehicles that do not carry human operators, use aerodynamic forces of lift, can fly autonomously or be piloted remotely, can be expendable or recoverable, and can carry lethal or non-lethal payloads (Blazakis, 2004). The role of UAVs in the military has rapidly expanded over the years such that every branch of the U.S. military deploys some form of UAV in their intelligence, surveillance, and reconnaissance operations. Recent U. S. military successes include a USAF Predator UAV operating in Iraq that successfully aided in finding Saddam Hussein (Rogers, 2004). Perhaps the most amazing fact from this is that the crew which was actively in control of the UAV, was located in Nellis AFB in Las Vegas, Nevada. Another more recent example took place in August 2004 when a Predator UAV armed with Hellfire missiles rescued a group of U. S. Marines pinned down by sniper fire in Najaf, Iraq. That Predator was also controlled from Nellis AFB in Las Vegas, Nevada. The worth of UAVs has become such that the militaries of every major power on the planet employs the use of UAVs including, but not limited to Germany, England, China, France, Canada, South Africa, and Israel.

The use of UAVs has also become so popular that many civilian uses have arisen, from security and law enforcement uses such as border and wildfire surveillance, to agricultural uses such as crop dusting and crop health monitoring. For example, the NASA ERAST Pathfinder has been successful in monitoring coffee fields in Hawaii for ripe beans, which has lowered operating costs and increased revenue for the company (Roeder, 2003). UAVs have been so successful, that future planned missions to Mars will see the use of UAVs to explore the Martian surface. Other uses for UAVs will eventually include communication relay and weather monitoring by high altitude-long endurance (HALE) platforms as well as surveillance and reconnaissance in the service of Homeland Defense.

UAV Mishaps

For all their successes and usefulness, the operational record of UAVs has been marred by high mishap rates which are frequently cited as a deterrent to the widespread use of UAVs. Mishaps as defined by the U. S. Navy, are unplanned events that directly involve naval aircraft, which results in $10,000 or greater cumulative damage to aircraft or personal injury. Under this classification, a "Class A" mishap is that in which the total amount of damage exceeds $1,000,000 or results in the destruction of the aircraft. The high mishap rate, which is currently 100 times higher than that of manned aircraft, has proved to be a deterrent to the military fully embracing the use of UAVs. For example, the Pioneer UAV has an unacceptable Class A mishap rate of 385 mishaps per 100,000 flight hours since 1986. In contrast, manned Naval aviation has a rate of 2 mishaps per 100,000 flight hours (Jackson, 2003). The Predator UAV, which has a total operational hour count of under 100,000 hours, has had 74 mishaps contrasted with a mishap rate of 8.1 per 100,000 flight hours for manned civil and commercial aircraft.

Schmidt & Parker (as cited in Fergusen, 1999), examined 107 mishaps that occurred between 1986 and 1993 and found that 59% were attributable to electromechanical failure and 33% were due to human errors attributed from crew selection and training, pilot proficiency, personnel shortages, operational tempo, and errors in teamwork and aircraft control. Seagle (as cited in Fergusen, 1999) also examined 203 mishaps from 1986 through 1997 and found that 43% of those were attributable to human error. One example of a mishap occurred when a Predator UAV encountered a fuel problem during a descent and upon entering instrument meteorological conditions, icing occurred and the engine lost power. The UAV crashed in an unpopulated wooded area so there were no casualties. It was determined that the operators' attention became too focused on flying the UAV in conditions they had rarely encountered. Ultimately, there was a lack of communication between the two operators during the emergency, which resulted in the mishap.

The increasing frequency and varied applications in which UAVs are being, and will be used, coupled with the high mishap rate speak to the need for more human factors research. There is much work to be done in many areas including automation, vigilance, feedback, procedures, crew selection, displays, training, coordination, and communication. Given today's emphasis on teamwork and the foreseeable future of UAV command and control possibly emphasizing teams of teams of UAVs working in concert in a heterogeneous network-centric battlefield, we have identified the coordination and command and control aspects of UAVs as a critical research issue.

Myths and Fallacies

Despite the apparent usefulness and worth of UAVs, and given their high mishap rate, very little human factors work in this area has been done. We believe that the lack of human factors work in the area is due to several myths and fallacies that surround the operation of UAVs. We feel that these false beliefs hide the fact that there is much research that is needed in this field. By shedding light on these fallacies, we hope to draw attention to the current human factors issues as well as any potential problems that might arise in future systems.

The Automation Fallacy

UAVs are highly automated. Platforms such as the Global Hawk are capable of taking off, flying missions, and landing, all fully autonomously. The belief is that more automation is better and if there is a problem, a person can simply step in and deal with it. However, over thirty years of sponsored research has shown that automation changes the human's task and not always in a positive manner. Many mishaps are attributed to the human being "out-of-the-loop," just as in manned aircraft such as commercial jetliners. We posit that one of the advantages of UAVs is that the humans have the ability to override the automation and perform dynamic re-tasking.

The Air Traffic Control Fallacy

Another fallacy concerns the belief that since air traffic controllers can monitor dozens of vehicles, UAV operators should also be able to handle multiple platforms at once. The fact here is UAV control tasks involve much more that monitoring and control of aircraft position. Many platforms such as the U. S. Army Shadow and the U. S. Navy Pioneer are controlled by stick and rudder controls. Dynamic re-tasking and re-planning maximally exploits the system. In addition, many believe that the state of the art is 1 operator per vehicle and that a 1:4 operator to vehicle ratio is a logical extension (Shope, DeJoode, Cooke, & Pedersen, 2004). However, the current state of practice demonstrates a 2:1 operator to platform ratio and current research suggests that a 1:n operator to UAV ratio will prove to be problematic.

The Manned Flight Fallacy

This fallacy stems from the belief that UAV flight is no different from manned flight. Since the UAV is a vehicle, piloting a UAV is similar to piloting an airplane in the cockpit, thus a single pilot should be sufficient. The truth is that a UAV is not simply a vehicle, but a system that includes ground control, operators, intelligence, weather personnel, maintenance personnel, and payload operators in addition to the UAV itself. This "piloting analogy" ignores years of studies on time lad, loss of visual cues, depth perception, and ignores the system functions that go beyond flight such as re-tasking, re-planning, and sensor operation.

The Unmanned Fallacy

That UAVs are unmanned, and even the name "unmanned," has propagated the myth that UAVs are indeed 'unmanned.' This notion could not be farther from the truth however as there are always humans in the loop at one point or another whether it is preprogramming a UAV to takeoff, fly a set of waypoints, and land autonomously, to the pilot that is actually controlling the UAV via stick and rudder controls. The fact that the UAV is uninhabited such that there is no actual flight crew onboard does not mean that it is unmanned. The two examples previously discussed above highlight the fact that even though the crews in control of the UAVs were roughly 7,000 miles away, there were nevertheless, humans involved in the loop. This "unmanned fallacy" assumes that since there are no humans in the loop, there is therefore, no need for human factors. However, data gathered from the examination of mishaps demonstrates that humans are indeed a part of UAV control and that human factors research should be an iterative part of the design and implementation of UAV systems as well the training of personnel and the development of operational procedures.

Principles of Command and Control

Advances in technology have increased the cognitive complexity of tasks and therefore, the need for teamwork has also increased. Teams operating in highly cognitive domains (e.g., aircraft cockpits, air traffic control, operating rooms) are required to plan, detect and interpret cues, make decisions, and perform as one coordinated unit. We define teams as a distinguishable set of two or more people who interact dynamically, interdependently, and adaptively toward a common and valued goal, who have each been assigned specific roles or functions to perform, and who have a limited life span of membership (Salas, Dickinson, Converse, & Tannenbaum, 1992). The collaborative cognitive processes that teams undergo are referred to as *team cognition*. Team cognition is more than the sum of the cognition of individual team members. Instead, it emerges from the interplay of the individual cognition of each team member and team process behaviors.

Why measure team cognition? Team cognition contributes to team performance now more than ever in today's cognitive tasks. Many organizations (military and civilian) hold the belief that teams are the solution to many problems. It is perceived that teams are better able to handle stress, are more adaptable in dynamic environments, make better decisions, and are more productive than individuals alone. Research on understanding team cognition and effective team performance has long been an area of intense focus for human factors, military, social, cognitive, and industrial/organizational psychologists (Cooke, Salas, Kiekel, & Bell, 2004).

Now, more than ever, issues of assessing team performance, training teams, and designing technological aids for effective team command-and-control performance are critical, yet highly challenging. How can team performance be measured? How can we characterize and assess cognitive skill at the team level? Can assessment occur without disruption of operational performance and can it occur in time for intervention? How is team cognition and performance impacted by training, technology, and team composition? Is team cognition different than the sum of the cognition of individual team members? What are effective training regimes or decision tools for these team members?

Our research program in the CERTT (Cognitive Engineering Research on Team Tasks) Laboratory is focused on these and other questions pertaining to team performance and cognition. Team coordination is characterized by timely and adaptive information exchange among team members. More specifically, command-and-control tasks in both military and civilian domains can be characterized as challenging from the perspective of the command-and-control team for a number of reasons including the; 1) unanticipated nature of the situation, 2) *ad hoc* formation of team structure, 3) lack of familiarity among team members, and 4) extended intervals with little or no team training. Items 3 and 4 are particularly relevant to military and civilian command-and control communities because there can be fairly long periods when command-and-control teams are not able to train and practice together, yet they are expected to be competent as soon as they are deployed. We view team coordination as central to team skill in command-and-control. In addition, for teams that stay together in a natural, operational setting (e.g., UAV teams) it is difficult to control the amount of exposure teams get to the operational tasks between laboratory sessions. Other goals of the CERTT Laboratory include the identification of issues and needs in the measurement of team cognition, the development and evaluation of new measures and the application of new measures and methods in which to better understand and evaluate team cognition.

The CERTT Laboratory

The heart of the CERTT Laboratory, shown in Figure 1, is a flexible synthetic task environment (STE) that is designed to study many different synthetic tasks for teams working on complex environments. STEs provide an ideal environment for study of team cognition in complex settings by providing a middle-ground between the highly artificial tasks commonly found in laboratories and the often uncontrollable conditions found in the field. We are currently studying team cognition with the use of an UAV-STE controlled by a three-person team whose mission is to take reconnaissance photographs. This current set-up is based on a cognitive task analysis of the ground control station of the Predator UAV operated by the U.S. Air Force (Gugerty, DeBoom, Walker, & Burns, 1999). The UAV-STE emphasizes many team aspects of tasks found in UAV operations such as planning, re-planning, decision-making, and coordination.

Figure 1. *CERTT Lab participant and experimenter consoles.*

The team members involved in this task are the Air Vehicle Operator (AVO) who flies the UAV by controlling the heading, altitude, and airspeed, the Payload Operator (PLO) who controls camera settings and takes reconnaissance photos, and the Data Exploitation, Mission Planning and Communications Operator (DEMPC) who plans the mission and acts as the navigator. More information on the CERTT Laboratory can be found in other publications (Cooke, Rivera, Shope, & Caukwell, 1999, Cooke & Shope, 2002).

Our Findings

Team Performance We use performance data as the criterion against which other measures (i.e., team process behaviors, taskwork knowledge, teamwork knowledge, situation awareness) can be evaluated. For instance, if one of our cognitive measures fails to predict performance differences, then it is not as useful as one that does. All interventions, personnel selection rules, manipulations, technological innovations, decision aids, or training strategies are of little importance if they have no impact on this bottom line. As a result, much of the team literature has focused on measures of team performance or effectiveness and findings that impact team performance or effectiveness (e.g., Salas et. al., 1992). In our UAV-STE, we rely on a composite measure of team performance that includes number of targets photographed, number of airspace violations, amount of consumables used (i.e. fuel, film), and time spent in alarm or warning state.

Thus far, we have completed 5 separate experiments which have examined team performance and cognition under varying circumstances including the co-location (all three members in the same room) vs. distribution (members located in different rooms) of team members, encouragement vs. discouragement of information sharing during breaks, and the "force-feeding" of teamwork and coordination information prior to the development of taskwork knowledge. Results from prior experiments indicate that the encouragement vs. discouragement of information sharing had no effect on team performance and that attempts to "force-feed" teamwork and coordination information were unsuccessful, suggesting a sequential dependence of knowledge development such that taskwork knowledge must precede teamwork knowledge. Our findings have also shown that geographic distribution of team members had no effect on performance. Distribution did however, have an effect on process behaviors and knowledge.

In addition to team performance, we measure process behavior in our UAV task through experimenter observations and ratings. Experimenters monitor behaviors such as communication, coordination, and leadership behaviors and rate them on a scale that indicates the observed quality of these behaviors. Also behavior is observed and rated at critical event junctures in the simulation. Overall, we find that process data can provide information where performance data do not. In some cases we find that outcome does not differ, but process does, providing some insight into the teams' adaptive behaviors. In the experiment described above, we found that co-located and distributed teams behaved differently, but managed to obtain similar performance scores (Cooke, et. al., 2004). Without the process data we might have assumed that there was no impact of distributed or co-located settings, but in conjunction with process data, we now understand that team interactions were adaptive for their own environment and the adaptation of the best teams may provide insight for training or design interventions.

Overall, the lack of performance effects is good news for military and civilian agencies which have begun to embrace distributed command-and-control. This is especially beneficial for the operation of teams-of-teams of UAV operators that must coordinate and work in concert, yet are geographically distributed. However, a caveat here is that teams need to be free to adapt their coordination behavior to preserve performance effectiveness. Thus, command-and-control environments and procedures demand careful consideration of these human factors issues.

Team Practice In our UAV task we have found consistent and robust findings in regard to team skill acquisition and in some cases, retention of that skill. Individuals are trained to criterion on the AVO, PLO, or DEMPC task prior to working together as a team. Once they come together in a mission scenario as a team, it takes them 3-4 40-minute missions to reach asymptotic levels of team performance (Figure 2).

Our knowledge measures indicate that most taskwork and teamwork knowledge is stable by the first mission. The process and communications data, on the other hand, indicate that teams during this initial period of working together are learning how to coordinate of pass information back and forth in a timely and adaptive manner. There is also a hint of loss due to a retention interval when some teams returned after several weeks for their third session (after Mission 7). The study of retention intervals on coordination skills is currently being tested in the laboratory.

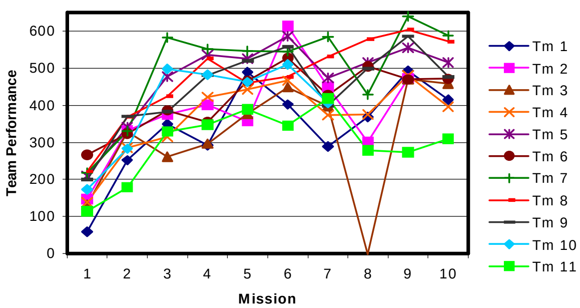


Figure 2. *Team performance of 11 teams over 10 missions. A long break occurred between Missions 7-8*

Team Communication Team communication is central in command and control tasks. Communication is also a critical mode by which coordination occurs, though it is possible to communicate in a variety of different ways (e.g., oral, gestural, computer messaging) and it is possible to coordinate implicitly without communication. In our UAV experiments we have found that communication patterns (both the content of what is said and the flow from person to person) are associated with team performance (Kiekel, Cooke, Foltz, & Shope, 2001).

Effective teams have different patterns compared to ineffective teams. Effective teams are generally more consistent in their communication patterns than ineffective teams. Workload influences patterns. Other subtle factors such as geographic distribution also influence communication patterns. Communication patterns change as teams acquire experience. Why are we interested in communication? It is not so much to train teams in ways to better communicate, thereby enhancing coordination, though that would be one approach. Rather we view communication as a readily available source of information on team cognition. Again, because we view team cognition as an emergent property of teams and believe that cognitive processing at the team level takes place in the interactions among team members, we see communication as a direct reflection of team cognition. Like team cognition, the communication-based measures should predict team performance, but should also provide additional diagnostic information. After having identified patterns associated with ineffective and effective teams, we are now exploring finer distinctions among teams in regard to team knowledge and team situation awareness that can be ascertained through analysis of communication data. We are also identifying was to automate this process with the ultimate goal of embedded and on-line communication analysis leading to a diagnosis of a team's cognitive state.

Implications for UAV Operations

The success of UAVs in both military and civilian applications is much more complex than is commonly thought as demonstrated by the various myths and fallacies that exist regarding their operation. The complexity of operations is also likely to become even higher as more UAVs take to the skies, flying longer, more varied missions. While this may not be as important an issue to military forces operating in sparsely populated areas of the world, this is of special concern in populated civilian areas. The Federal Aviation Administration (FAA) has mandated that in order for UAVs to operate in the national airspace (NAS), certain safety issues must be addressed. These issues include the need for collision avoidance, and over-the-horizon subsystems development, leading up to the establishment of certification processes and operating criteria (National Aeronautics and Space Administration, 2001). These concerns stem from the simple fact UAV operation in the NAS is hazardous because there is no pilot onboard that can aviate, navigate, communicate, diagnose problems, and scan the environment for traffic.

Despite the inevitable advances in collision avoidance and over-the-horizon technologies, chances of mishaps will still become higher due to the increased traffic and coordination requirements on teams of UAVs. Coupled with the aspect of UAV operators working in teams of teams, controlling multiple platforms, and interacting with manned air traffic and air traffic controllers, the need for interventions stemming from the study of performance, training, communication and coordination in UAV operations will become a valuable commodity. In addition, our research has shown that the coordination among only 3 ground control personnel controlling a single UAV is highly complex. Studies have yet to be conducted in the coordination of all personnel (i.e. operators, maintenance staff, air traffic control) involved in the operation of a single system. In addition, future military doctrine calls for an increase in the UAV to operator ratio where it is thought that one operator will control multiple UAVs. What will be the impact on coordination? What will happen when single operators controlling multiple UAVs must coordinate and interact with other operators performing the same task, air traffic control, and other manned aircraft?

It is the goal of the CERTT Laboratory to explore team coordination and in the process, dispel the myths and fallacies that reside within UAV operations. Raising the awareness of the myths and issues involved in UAV operations within the human factors community will increase the amount of research done in this budding area. Such research will not only benefit UAV operators, but will answer questions such as those above, as well as increase the safety of air operations in both military and civilian sectors.

References

Blazakis, J. (2004). *Border security and unmanned vehicles* (CRS Rep. No. RS21698). Washington D. C.: Congressional Research Service.

Cooke, N., DeJoode, J., Pedersen, H., Gorman, J., Connor, O., & Kiekel, P. (2004). *The role of individual and team cognition in uninhabited air vehicle command-and-control* (Final Performance Report). Mesa, AZ, Arizona State University East.

Cooke, N. J., Rivera, K., Shope, S. M., & Caukwell, S. (1999). A synthetic task environment for team cognition research. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting,* 303-307.

Cooke, N. J., Salas, E., Kiekel, P. A., & Bell, B. (2004). Advances in measuring team cognition. In E. Salas and S. M. Fiore (Eds.), *Team Cognition: Understanding the Factors that Drive Process and Performance, pp. 83-106,* Washington, DC: American Psychological Association.

Cooke, N. J. & Shope, S. M. (2002). The CERTT-UAV Task: A Synthetic Task Environment to Facilitate Team Research. *Proceedings of the Advanced Simulation Technologies Conference: Military, Government, and Aerospace Simulation Symposium*, pp. 25-30. San Diego, CA: The Society for Modeling and Simulation International.

Fergusen, M. G. (1999). *Stochastic modeling of naval unmanned aerial vehicle mishaps: Assessment of potential intervention strategies*. Unpublished master's thesis, Naval Post Graduate School, Monterey.

Gugerty, L., DeBoom, D., Walker, R., & Burns, J. (1999). Developing a simulated uninhabited aerial vehicle (UAV) task based on cognitive task analysis: Task analysis results and preliminary simulator data. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp.86-90). Santa Monica, CA: Human Factors and Ergonomics Society.

Kiekel, P. A., Cooke, N. J., Foltz, P. W., & Shope, S. M. (2001). Automating measurement of team cognition through analysis of communication data. In M. J. Smith, G. Salvendy, D. Harris, and R. J. Koubek (Eds.), *Usability Evaluation and Interface Design*, pp. 1382-1386, Mahwah, NJ: Lawrence Erlbaum Associates.

Jackson, P. (2003). All the world's aircraft. Alexandria, VA: Janes Information Group, pp. 721-722.

NASA ERAST Certification and regulatory roadmap (2001). Retrieved January 25, 2005, from http://www.psl.nmsu.edu/uav/roadmap/

Roeder, L. (2003, January-February). Hope from on high. *Unmanned Vehicles, 8(1),* 14-15.

Rogers, K. (2004, March 2). Nellis crew helped nab Saddam. *Las Vegas Review-Journal,* pp. 1A.

Salas, E., Dickinson, T., Converse, S., & Tannenbaum, S. (1992). Toward an understanding of team performance and training. In R. Swezey & E. Salas (Eds.). *Teams: Their Training and Performance* (pp. 3-30). Norwood NJ: Ablex Publishing Corp.

Shope, S., DeJoode, J., Cooke, N., & Pedersen, H. (2004). *Command and control interfaces for Predator squadron operations centers* (Final Technical Report). Mesa, AZ: U. S. Positioning.

MEASUREMENT OF SITUATION AWARENESS EFFECTS OF ADAPTIVE AUTOMATION OF AIR TRAFFIC CONTROL INFORMATION PROCESSING FUNCTIONS

Carlene M. Perry
Noa Segall
David B. Kaber

Department of Industrial Engineering
North Carolina State University
Raleigh, North Carolina

The goal of this research was to define a measure of situation awareness (SA) in an air traffic control (ATC) task and to investigate the effect of adaptive automation (AA) of various information processing functions on SA. An ATC simulation was used that was capable of presenting four different modes of control, including information acquisition, information analysis, decision making and action implementation automation, and a manual mode. Eight subjects completed two trials under each mode of control. Operator workload, assessed using a secondary task, was used to trigger automation of the primary ATC task. The SA measure was an adaptation of the Situation Awareness Global Assessment Technique (SAGAT), involving cueing of aircraft positions as well as objective weighting of the relevance of aircraft to controllers for queries. The SA response measure revealed a significant effect of AA on subject perception and overall SA, with superior SA under the information acquisition mode of automation. ATC performance was significantly superior ($p<0.05$) when automation was applied to lower-order sensory processing functions, including information acquisition and action implementation, as compared to higher-order functions, specifically information analysis. During manual control periods as part of AA trials, ATC performance was significantly superior when following automation of information acquisition and information analysis functions. Secondary task performance was significantly worse under information analysis and decision making automation.

Introduction

Air traffic control (ATC) requires high levels of cognitive processing, and one approach for alleviating stress and workload of controllers is to use automation to perform some controller activities (National Research Council (NRC), 1998; Parasuraman & Riley, 1997). Automation has many potential advantages for controllers, including reduced task load (Laois & Giannacourou, 1995) and increased system reliability (NRC, 1998). However, automation in ATC may also present disadvantages (Dillingham, 1998), including a loss of controller situation awareness (SA) (Endsley & Jones, 1995). As machines perform more and more ATC functions, controllers have less interaction with the traffic management system, impairing their ability to detect when a problem has occurred, determine the current state of the system, understand what has happened and what courses of action are needed, and react to the situation (Endsley, 1996). Thus, maintaining SA in ATC is critical for accurate decision making and performance (Endsley, 1996), and this issue needs to be addressed through automation design.

Currently, advanced forms of adaptive automation (AA) are being considered for ATC to mitigate out-of-the-loop (OOTL) performance problems associated with conventional automation, and to preserve operator SA. AA refers to complex systems in which the level of automation or the number of system functions being automated can be modified in real time (Scerbo, 1996). Some research has explored the use of dynamic function allocations (DFAs) in the context of ATC simulations. Results provide evidence that AA may improve ATC performance over completely manual control and static automation (e.g., Hilburn, Jorna, Byrne & Parasuraman, 1997). They also indicate that the effectiveness of AA in the context of ATC may be dependent upon the type of automation presented to an operator. For example, Clamann, Wright, and Kaber (2002) found that, in the context of a low-fidelity ATC simulation, humans are better able to adapt to AA (from a performance perspective) when applied to lower-order sensory and psychomotor functions, such as information acquisition and action implementation, as compared to AA applied to cognitive (planning and decision making) tasks.

Measures of SA and AA

Many measures of SA have been developed over the past 10-15 years, including direct, objective measures such as the Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1995). SAGAT involves comparing an operator's perceptions of a task environment to some "ground truth" reality. This is accomplished by freezing a simulation exercise at random points in time and hiding task information sources (e.g., blanking visual displays) while subjects quickly answer questions about their current knowledge of the simulation. Subject responses are then graded based on actual data on the real situation, thus providing an objective measure of SA.

AA research has demonstrated SAGAT to be sensitive to dynamic changes in system states (Kaber & Endsley, 2004), as well as changes in adaptive interface content over time (Kaber & Wright, 2003). Kaber and Endsley (2004) also observed that operators achieved better SA with DFAs of levels of automation that applied computer assistance to decision-making aspects of the dynamic control task, as compared to levels applying automation to monitoring and implementation roles. This research also suggests that the impact of AA on SA may be dependent upon the human-machine system information processing (IP) functions to which AA is applied, but that SA may be affected in a different way than performance.

Recent research examining the use of SAGAT in an ATC simulation (McClernon, 2003) found its sensitivity for identifying differences among manual and automated conditions to be limited. Nunes (2003), who applied SAGAT to evaluate aided and unaided display conditions in an ATC task, also found that the technique did not reveal differences between conditions. Another study by Hauss and Eyferth (2003) suggests that SAGAT may not be a sensitive measure of SA in the ATC environment due to different aircraft having different relevance to controllers at different times. They argued that aircraft which had recently been contacted by a controller, or required control actions, demanded more attentional resources than other displayed aircraft. Consequently, controllers may focus on certain aircraft to the exclusion of others at various times and may recall their flight parameters in responding to SAGAT queries more accurately.

Hauss and Eyferth's (2003) concerns with the SAGAT for assessing controller SA led them to develop a new measure of SA which assigned weights to aircraft based on their relevance to the current control scenario. In addition, rather than having subjects recall aircraft positions on a blank radarscope, as an initial query, they employed cued recall in which participants were given the positions for the aircraft they were to be queried on. Hauss and Eyferth (2003) compared the new SA measure with SAGAT using a high-fidelity air traffic management simulation. Their results confirmed controllers used event-based mental representations, since significantly more relevant parameters than irrelevant parameters were reproduced using the new measure.

In the current study, we developed a modified approach to implementation of the SAGAT measure in order to assess the impact of various forms of AA of ATC IP functions on SA. Cued recall of aircraft positions in a simulated ATC task was implemented and aircraft relevance was objectively weighted as a basis for SAGAT queries. Different from Hauss and Eyferth's (2003) measure, this approach involved real-time identification of aircraft in conflict, as well as those that had recently been issued clearances (e.g., hold, reduce speed, etc.), as predictors of aircraft relevance to controllers. It was expected that these modifications would lead to a more sensitive measure of the impact of AA on controller SA, as compared to the SAGAT measures implemented by McClernon (2003) and Nunes (2003).

Method

We evaluated the SA, performance, and workload effects of AA of four different stages of IP in ATC. The forms of automation included information acquisition, information analysis, decision making and action implementation (see Parasuraman, Sheridan, & Wickens, 2000).

Tasks

The Multitask© Simulation. Multitask© is a lab simulation of ATC developed for studies of workload-matched AA of ATC IP functions (see, for example, Clamann et al., 2002). The task interface (see Figure 1) includes a radarscope, control and status boxes and a menu bar. Near the center of the radarscope are two airports. Each airport has two runways. Eight equally spaced holding fixes are also represented on the display by small circles (approximately 30 nm from the airports).

Simulated aircraft are represented on the display by triangle icons and data tags presenting their call signs. The aircraft icons represent one of three possible aircraft types: commercial, private, or military. The type of aircraft also dictates the possible range of speeds for the vehicle. During simulation run time, aircraft first appear toward the perimeter of the display on one of eight approach trajectories and move toward one of the two airports, destined for one of the two runways at an airport.

The control box includes eight buttons. Five control buttons facilitate clearance change commands, including reduce speed, hold, resume, change airport, and change runway. Two action commands are used to submit and cancel these clearances. Finally, the query command is used to initiate communication with an aircraft and obtain its flight parameters.

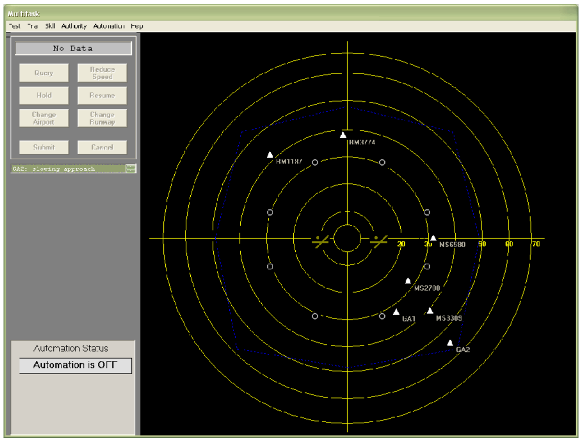


Figure 1. *Multitask© display in manual control.*

The simulation is capable of operating under one of the following five modes of automation:

(1) Manual control – No assistance is provided.

(2) Information acquisition – A scan line rotates around the radar display, and as it passes over an aircraft icon, a Trajectory Projection Aid (TPA) for that aircraft is presented for 2 sec. The TPA shows the aircraft destination and route, as well as its speed and destination airport and runway identifiers. This form of automation assists operators with acquisition of data on aircraft.

(3) Information analysis – Information on each aircraft on the radarscope is displayed in a table, including the aircraft's call sign, destination airport, destination runway, speed, and distance from the airport. A final column denotes the call sign of aircraft that are in conflict with each other. This form of automation assists operators with the integration of aircraft information.

(4) Decision making – In addition to conflict alerting, recommendations for conflict resolution are provided. Information on conflicting aircraft, the recommended clearance change, and which aircraft to advise of the change, are all displayed. This form of automation assists operators with IP requirements associated with decision and response selection aspects of the task.

(5) Action implementation – This form of automation simulates the "hand-off" of aircraft control from approach control to local-tower control, and the tower automatically maintains full control responsibility for aircraft within 20 nm of the center of the radarscope. This type of automation prevents any conflicts after "hand-off" to tower control. Action implementation automation assists the operator with the requirement of response execution as part of the ATC simulation.

Under all modes of automation, the objectives of the controller are to contact aircraft appearing on the radar display and make any necessary changes to pre-existing aircraft clearances (based on their potential to cause a conflict) while maintaining landing efficiency. Multitask© performance is measured in terms of the number of aircraft cleared, the number of trajectory conflicts, and actual collisions. This data is recorded during simulation trials. Aircraft arriving safely at an airport are considered cleared. Aircraft traveling within 3 nm of other aircraft, or two aircraft that are within 20 nm of the center of the radarscope and destined for the same runway at the same airport, are considered to be in conflict. Aircraft that simultaneously arrive at the same airport destined for the same runway, or aircraft that come in contact with each other, constitute actual collisions. During experimental trials, the various modes of automated assistance can be switched "on" or "off", based on operator workload states; however, only one mode can be used per trial.

Secondary Gauge-Monitoring Task. The experiment used a dual-task scenario involving simultaneous subject performance of the Multitask© simulation and a gauge monitoring task to objectively assess operator workload. The gauge task included a fixed scale, moving pointer display with a central "acceptable" region bordered on either side by two "unacceptable" regions. The user's goal was to detect and correct pointer deviations into either unacceptable region by using a keyboard. Gauge task performance was recorded as a hit-to-signal ratio.

Experimental Design and Procedures

Approach to AA. The gauge-monitoring task provided an index of operator workload in the Multitask© simulation. A low score in the gauge task implied a high level of workload in the ATC simulation and vice versa. The gauge task served as a basis for triggering DFAs in Multitask©. When secondary-task performance was poor, suggesting an increase in operator workload, the ATC simulation shifted from manual control to automated control. If operator secondary-task performance was good, the simulation returned to manual control.

Experiment Design. The experiment followed a within-subjects design with blocking on the subject. Eight subjects completed two trials under each of the five modes of Multitask© control. Each trial lasted approximately 50 minutes, including 30 minutes of simulation time and approximately 20 minutes to answer SA questions.

Situation Awareness Measure. The modified SAGAT measure (described above) was developed based on a goal-directed task analysis (GDTA) of ATC operations and application of the GDTA methodology to the Multitask© simulation. Following Endsley's (1995) methodology, three simulation freezes were conducted at random points in time during experimental trials to deliver SA queries. Subjects were posed with 9 questions during each freeze, including three targeting each level of SA (1 – perception; 2 – comprehension; 3 – projection), as defined by Endsley (1995). When a freeze occurred, the simulation displays were temporarily blanked and subjects were asked to move to a secondary computer workstation and respond to queries. At the same time, an experimenter collected information from the Multitask© software by accessing an automated aid which provided information on aircraft in conflict with each other and recommended clearances. Based on this information, the experimenter identified the three aircraft with the highest priority, or greatest "relevance", at that point in time in the simulation. Aircraft priority was determined based on a hierarchy of simulation events, e.g. aircraft in conflict were considered to have the highest priority, followed by aircraft issued a "hold" clearance, etc. The experimenter then sketched the locations of the "high priority" aircraft on a blank graphic of the Multitask© radarscope. The subjects were given the graphic and asked to respond to each of the 9 SA queries for each "high priority" aircraft. Composite scores for Level 1, 2, and 3 SA were computed based on the accuracy of subject responses to the sets of questions across freezes.

Hypotheses

(H1) We expected the modified SAGAT measure to be sensitive to changes in controller SA as a result of the AA manipulations. Counter to Kaber and Endsley's (2004) findings, because of the complexity of the version of Multitask© used in this study, subjects were expected to do better at responding to SA queries under lower levels of automation (information acquisition) and manual control as compared to high-level automation (information analysis and decision making), as a result of the potential for OOTL performance problems. We also speculated that under high levels of automation, such as decision making or information analysis, operators would exploit the additional capabilities of the automation, including conflict warnings and recommendations, pay less attention to the actual radarscope, and spend less time on low-level control functions which may be important to achieving SA.

(H2) On the basis of Clamann et al. (2002) findings, we expected Multitask© performance to be superior during trials in which AA was applied to lower-order sensory/ response functions, such as information acquisition and action implementation.

(H3) Based on Hilburn et al. (1997) results, AA of Multitask© was expected to affect performance on the secondary gauge-monitoring task, or operator workload. It was expected that higher levels of automation, including information analysis and decision making, presenting complex displays for operator interpretation, might demand high levels of visual attention and increase workload.

Results and Discussion

Situation Awareness

An ANOVA on the SA response measures revealed a significant effect of the specific forms of AA on Level 1 SA queries (F(4,227)=3.78, p=0.005) and the total SA score (F(4,227)=2.7, p=0.032). These findings support our expectation (H1) that the modified version of the SAGAT-based measure was sensitive to AA manipulations. Figure 2 shows the average Level 1 SA scores under each mode of automation. The pattern of results on Total SA was similar.

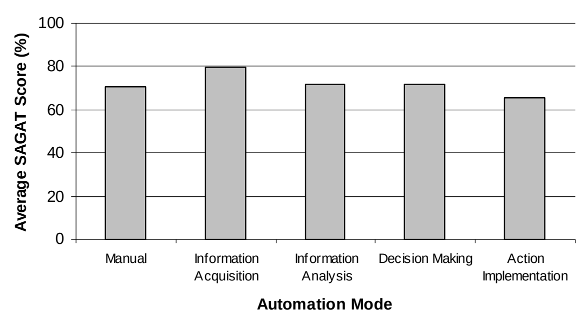


Figure 2. *Mean Level 1 SAGAT scores for the different modes of automation.*

Duncan's test showed Level 1 SA to be significantly superior under information acquisition automation, compared to information analysis, decision making, action implementation, and manual trials (p<0.05). However, manual control was not found to increase Level 1 SA. With respect to total SA, Duncan's test also revealed information acquisition to be superior to action implementation automation (p<0.05), which is consistent with our hypothesis (H1). However, SA during information analysis and decision making trials was not inferior to SA during other automation trials. In general, these results suggest that perception of system states may be most critically affected by demanding automation displays.

Figure 3 summarizes the mean Level 1, Level 2, Level 3 and total SA scores for automated and manual control periods, as part of AA trials (only). A marginally significant effect of the mode of automation was found for Level 2 SA queries (F(1,227)=3.51, p=0.062), indicating that subject comprehension was, on average, higher during manual control periods compared to automated control periods. This finding supports the notion that introducing some forms of automation in ATC may remove the controller from the loop (Endsley, 1996) and lead to decrements in higher levels of SA (H1).

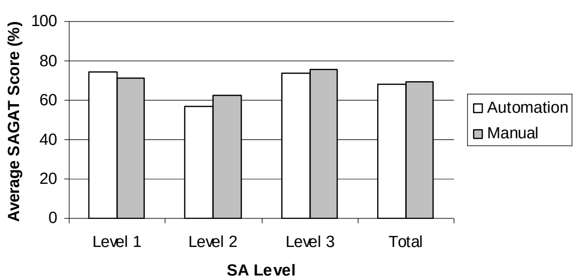


Figure 3. *Mean SAGAT scores during manual and automation control periods.*

Primary Task Performance

Results of ANOVAs on data collected during the automated control periods as part of AA revealed a significant effect of mode of automation on the number of cleared aircraft (F(3,41)=3.62, p=0.021) and the number of aircraft conflicts (F(3,41)=3.97 p=0.014), but not on the number of collisions (Figure 4). Duncan's MR test indicated that the number of cleared aircraft was higher for the information acquisition, decision making, and action implementation modes of automation, as compared to information analysis (p<0.05), in support of our hypothesis (H2). The high number of cleared aircraft during decision making may be attributable to the longer automated control periods under this mode of automation, as compared with the other modes. Duncan's test also revealed decision making to be significantly worse than information analysis for preventing aircraft conflicts (p<0.05). This finding was not surprising given that the decision aid made recommendations to subjects for dealing with conflicts. It is possible that subjects developed a strategy of waiting for the automation to warn them of a conflict and then to think about how to appropriately clear aircraft.

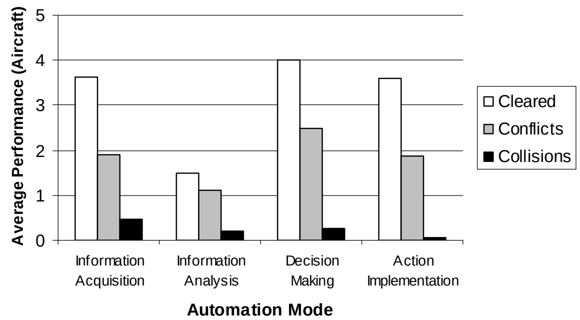


Figure 4. *Primary task performance during automated control periods.*

ANOVA results on manual control periods as part of the AA conditions revealed a significant effect of mode of automation on only the number of cleared aircraft (F(4,68)=7.58, p<0.0001). Safe landings were significantly higher for the information acquisition and information analysis modes of automation than for decision making and action implementation (Duncan's test, p<0.05). The results on the decision making condition are in agreement with our hypothesis (H2). It is possible that participants needed more time to shift from using a complex mental model for interaction with the decision aid back to their manual control mental model after the decision aid disappeared from the display and they had to identify conflicts themselves.

Workload (Secondary Task Performance)

An ANOVA on the workload data revealed a significant mode of automation effect when analyzing the automated control periods as part of AA trials (F(3,41)=4.01, p=0.014). Duncan's MR tests showed that action implementation, a lower-order sensory/response function, yielded higher average secondary-task performance than information analysis and decision-making automation (p<0.05). These findings are in line with our hypothesis (H3).

An ANOVA on workload data comparing the manual control condition with the manual control periods as part of AA also revealed a significant effect of the control mode (F(4,68)=2.66, p=0.04). The pattern of results under the manual control periods was almost exactly opposite to that observed during automated control periods. Duncan's test indicated that average workload was significantly lower (p<0.05) under decision-making automation, as compared to workload during manual control periods in AA of the information acquisition and action implementation functions, as well as the completely manual control condition. It is possible that when decision-making AA was applied and the recommendations for conflict avoidance were followed, the result was a lower workload when the simulation returned to manual control.

Conclusions

We designed a modified SAGAT approach to measuring SA in the context of an ATC task, which proved to be effective in terms of assessing the impact of specific forms of AA on controller perception, comprehension and projection. Using queued recall of aircraft, and establishing relevance weights for various aircraft at the time of SAGAT freezes, caused the SA response measures to be sensitive to the AA of information acquisition, information analysis, decision making, and action implementation functions. In general, our findings support a dependence of SA, performance, and workload effects of AA in ATC on the specific controller IP functions to which automation is applied. With a more complete understanding of the effects of AA on SA, additional research is needed to develop methods for real-time assessment of SA in ATC. Such a method could be used as a basis for triggering DFAs in complex, adaptive systems control on the basis of SA.

Acknowledgments

This research was supported by a grant from NASA Langley Research Center, Human Measures & Performance Element of the Airspace Systems Program, Hampton, VA (Grant No. NAG-1-03022). Lance Prinzel was the technical monitor. The views and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of NASA.

References

Clamann, M.P., Wright, M.C., & Kaber, D.B. (2002). Comparison of performance effects of adaptive automation applied to various stages of human-machine system information processing. In *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 342-346). Santa Monica, CA: Human Factors and Ergonomics Society.

Dillingham, G.L. (1998). *Air traffic control: Evolution and status of FAA's automation program.* (Tech. Report GAO/T-RCED/AIMD-98-85). Washington, DC: United States General Accounting Office.

Endsley, M.R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65-84.

Endsley, M.R. (1996). Automation and situation awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 163-181). Mahwah, NJ: Lawrence Erlbaum.

Endsley, M.R., & Jones, D.G. (1995). *Situation awareness requirements analysis for TRACON air traffic control* (Tech. Report TTU-IE-95-01). Atlantic City, NJ: Technology Center, Federal Aviation Administration.

Hauss, Y. & Eyferth, K. (2003). Securing future ATM-concepts' safety by measuring situation awareness in ATC. *Aerospace Science and Technology, 7,* 417-427.

Hilburn, B.G., Jorna, P., Byrne, E.A., & Parasuraman, R. (1997). The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload. In M. Mouloua & J. M. Koonce (Eds.), *Human-automation interaction: Research and practice* (pp. 84-91). Mahwah, NJ: Lawrence Erlbaum Assoc., Inc.

Kaber, D.B. & Endsley, M.R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5*(2), 113-153.

Kaber, D.B. & Wright, M.C. (2003). Automation-state changes and sensory cueing in telerobot control. In *Proceedings of the XVth Triennial Congress of the International Ergonomics Association* (CD-ROM). Seoul, Korea, August 24-29, 2003.

Laois, L. & Giannacourou, M. (1995). Perceived effects of advanced ATC functions on human activities: Results of a survey on controllers and experts. In *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 392-397). Columbus, OH: The Ohio State University.

McClernon, C.K. (2003). *Situation awareness effects of adaptive automation of various air traffic control information processing functions.* Unpublished masters thesis, North Carolina State University, Raleigh, NC, USA.

National Research Council (1998). *The future of air traffic control: Human operators and automation*. Washington, DC: National Academy Press.

Nunes, A. (2003). The impact of automation use on the mental model: Findings from the air traffic control domain. In *Proceedings of the Human Factors & Ergonomics Society 47th Annual Meeting* (pp. 66-70). Santa Monica, CA: Human Factors and Ergonomics Society.

Parasuraman, R., & Riley, V.A. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230-253.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model of types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics, 30*, 286-297.

Scerbo, M.W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 37-63). Mahwah, NJ: Lawrence Erlbaum.

THE GOOD, THE NOT-SO-BAD, AND THE UGLY: COMPUTER-DETECTED ALTITUDE, HEADING, AND SPEED CHANGES

Elaine M. Pfleiderer
FAA Civil Aerospace Medical Institute

The relationship between communication events and controller workload has been well established. Unfortunately, a substantial amount of time and effort is required to transcribe and code these events. Alternative measures might be preferable if they could be obtained more easily. Manning, Mills, Fox, Pfleiderer, and Mogilka (2002) found that, relative to a set of computer-derived measures, communication events might not add enough unique information to the prediction of subjective workload to justify the effort involved in obtaining them. At the time the study was conducted, computer-derived measures of altitude, heading, and speed changes were not available. The present investigation compares altitude, heading, and speed clearances with computer-derived measures of altitude, heading, and speed changes. Two 20-minute samples of live air traffic data were collected from each of four sectors in the Kansas City en route airspace. Communications data were transcribed from audio recordings and coded (e.g., altitude, heading, and speed clearances). Altitude, heading, and speed changes for each of the 20-minute samples were computed using the Performance and Objective Workload Evaluation Research (POWER) software system. The 20-minute samples were parsed into 4-minute intervals, and the number of communications events and changes were tallied for each interval. In addition, 16 subject-matter experts provided Air Traffic Workload Input Technique (ATWIT) measures for each 4-minute interval for all samples. Multiple regression analysis of altitude, heading, and speed clearances on mean ATWIT scores yielded an $R = .59$ ($R^2 = .35$). Multiple regression of the number of computer-detected altitude, heading, and speed changes on mean ATWIT scores yielded the same results. Multiple regression of both clearances and changes, employed to examine shared and unique variance of the two sets of measures, revealed that altitude changes alone could account for most of the variance in ATWIT scores ($R = .67$; $R^2 = .44$). Results suggest that computer-derived measures of altitude and heading changes may be a viable substitute for more labor-intensive communication measures. However, ground speeds recorded by the Host computer (and displayed on the controllers' radarscope) were too erratic to provide a valid measure of speed changes and could not be recommended as an acceptable alternative for speed clearances.

Introduction

Previous studies have shown that there is a relationship between communication events and controller workload (e.g., Bruce, 1993; Cardosi, 1993; Corker, Gore, Fleming & Lane, 2000; Morrow & Rodvold, 1998; Porterfield, 1997). However, it takes a considerable amount of time and effort to obtain these measures. First, audio recordings of all pilot and controller transmissions must be transcribed. Then, the transcriptions must be coded and the codes must be transferred to data files. To date, a satisfactory method for automating this process has not been developed. Consequently, the collection of communication events is time-consuming, labor-intensive, and subject to human error.

It would be preferable to develop alternative measures that might be more easily obtained. One possibility is the use of computer-detected changes in aircraft altitude, heading, and speed. Aircraft position information is routinely recorded by the en route Host computer system. Extraction of this information is fully automated, requiring minimal time and effort, and the resulting measures are completely objective.
In spite of the advantages of computer-detected measures, there is some question as to whether or not it is possible to develop algorithms sufficient to distinguish between random variability in aircraft position and actual changes. Computers rely solely on the adequacy of the parameters that have been set for them, and are generally unable to infer intent from partial information. With respect to aircraft changes, this inability may be problematic when determining if an actual change has occurred. Therefore, the first question to be answered has to do with the accuracy of computer-detected changes.

The second question to be answered has to do with the relationship between aircraft changes and controller clearances. We know that they do not necessarily share a simple "stimulus-response" relationship. Each has a separate set of associated workload factors. Prior to making a clearance, the controller must scan the airspace and make an assessment of the control situation. When the controller determines that a control action is necessary, then the decision must be made as to what particular control action will best fit the situation. Finally, there is the physical workload of issuing instructions to the pilot. In some instances, this includes making associated keyboard entries. After the clearance has been issued, another set of factors unfolds. First, the controller must make sure the clearance has been verbally acknowledged by the pilot and that the read back was accurate. The controller must monitor if the changes are being made in a timely manner and in accordance with the clearance. The controller must then evaluate whether the issued clearance was sufficient to produce the desired results or additional clearances will be necessary. Then there is the separate issue of increased traffic complexity produced by the clearance, not to mention the workload associated with changes in aircraft position in the absence of any clearance. So the question remains, given the delay and differences between them, how well can changes in aircraft position capture the "echo" of the workload associated with controller clearances?

The present study examined these two questions using the Performance and Objective Workload Evaluation Research (POWER; Mills, Pfleiderer, & Manning, 2002) software system, transcripts of controller clearances, and subjective measures of controller workload. The first phase of the study compared controller clearances with computer-detected changes. The second employed linear multiple regression to examine the relationship between controller clearances, computer-detected changes, and subjective measures of air traffic controller workload.

Method

Traffic Samples

Traffic samples were derived from National Airspace System (NAS) System Analysis Recordings (SARs) using National Track Analysis Program (NTAP) reports to obtain recorded altitude, heading, and ground speed information for each aircraft in the sample. Information in the text-based NTAP reports was organized into Microsoft Access database files using the NAS Data Management System (NDMS) program. (For a description of these programs and their output, see Mills, Pfleiderer, & Manning, 2002.) Traffic samples were obtained from four sectors in the Kansas City (ZKC) Air Route Traffic Control Center (ARTCC). Two 20-minute samples were taken from each of the four sectors. As shown in Table 1, two were high-altitude sectors and two were low-altitude sectors.

For testing the concordance of controller clearances with computer-detected changes, the samples were processed using the diagnostic log option of the POWER program. POWER diagnostic output includes the type of change detected (i.e., altitude, heading, or speed); the start and stop time of the change; the recorded altitude, heading, or speed at the beginning and at the end of the change; the duration and amount of change. For comparison with subjective measures of workload, the samples were POWER processed in 4-minute intervals, producing a total of 40 summary observations (i.e., the number of altitude, heading, and speed changes) for each measure.

Controller Clearances

Table 1. *Kansas City Air Route Traffic Control Center (ARTCC) Samples*

Sector/ Sample	Strata	Date	Sample Time (Local)
14A	High	01-20-99	07:16 - 07:36
14B	High	01-21-99	07:16 - 07:36
30A	High	01-21-99	09:40 - 10:00
30B	High	01-21-99	18:05 - 18:25
52A	Low	01-21-99	18:07 - 18:27
52B	Low	01-22-99	18:07 - 18:27
54A	Low	01-21-99	15:30 - 15:50
54B	Low	01-21-99	17:10 - 17:30

Controller clearances were obtained from voice tapes associated with the eight ZKC traffic samples. Time-stamped audiotapes of pilot and controller transmissions were transcribed and coded for content. All controller transmissions pertaining to altitude, heading, and speed clearances were extracted to construct a database containing transmission start time, transmission stop time, message content, and the type of clearance issued. Single transmissions containing more than one type of clearance were parsed into multiple records. For example, "COMAIR SIX TWENTY FIVE, FLY HEADING TWO EIGHT ZERO, MAINTAIN TWO FIVE ZERO KNOTS, AND DESCEND AND MAINTAIN FLIGHT LEVEL TWO ONE ZERO" would be represented by three separate records – one coded as a heading clearance (FLY HEADING TWO EIGHT ZERO), the second as a speed clearance (MAINTAIN TWO FIVE ZERO KNOTS), and the third as an altitude clearance (MAINTAIN TWO FIVE ZERO KNOTS). For comparison with subjective measures of workload, the clearances were tabulated in 4-minute intervals, producing a total of 40 summary observations.

Subjective Workload Measures

Subjective workload measures were contributed by 16 en route air traffic control instructors from the FAA Academy in Oklahoma City. All had formerly been Certified Professional Controllers (CPCs) at various en route centers across the United States. The participants received airspace training for each of the four sectors included in the traffic sample and then observed SATORI (Systematic Air Traffic Operations Research

Initiative; Rogers & Duke, 1993) re-creations of the live air traffic data. SATORI synchronizes extracted SAR data with voice tapes to produce graphical re-creations of air traffic events.

Participants provided subjective workload estimates using the Air Traffic Workload Input Technique (ATWIT; Stein, 1985). The ATWIT measures mental workload in "real-time" by presenting auditory and visual cues that prompt the participant to press one of seven buttons within a specified amount of time to indicate the level of mental workload experienced at that moment. Participants were prompted every four minutes during each traffic sample to provide an estimate of the amount of subjective workload they thought the radar controller responsible for the sector was experiencing at the time of the prompt. These assessments were summarized to produce a total of 40 mean subjective workload estimates.

Computer-detected Changes

Preliminary parameters for change detection were based on the *Private Pilot – Airplane Practical Test Standards* (FAA, 1995) that establishes guidelines for pilots regarding acceptable variability in altitude, heading, and speed. This seemed the best place to begin, since deviations beyond "acceptable variability" suggest that the aircraft was, in fact, responding to a clearance. Because Host computer system "glitches" sometimes occur in the recording of altitude, heading, and speed information (e.g., missing values recorded as an altitude of zero) an outlier criterion was established to ensure they would not be recorded as actual changes. A total of 900 individual flights (i.e., 300 flights for each type of change) were evaluated to determine the ability of the algorithms to detect altitude, heading, and speed changes. Accuracy of the computer-detected changes was tested by visual examination of graphs of each aircraft's altitude, heading, or speed that had been color-coded to highlight change parameters. Initial parameters were adjusted based on these evaluations Final parameters used in the analyses were as follows:

Altitude changes: Altitude must increase or decrease by a minimum of 200 feet per 12-second radar update and must continue to change in the same direction for at least 36 seconds. Changes of 10,000 feet or more occurring within a single update were automatically excluded as outliers.

Heading changes: Turns must be in excess of 10° per 12-second radar update and must continue in the same direction for at least 36 seconds. Changes greater than or equal to 40° occurring in a single update were automatically excluded as outliers.

Speed changes: Due to the extreme variability of ground speed, the algorithm for speed changes includes a running average calculated from all updates within each minute of data. If a cumulative change of 15 knots occurred within that time, individual 12-second radar updates were examined to determine the exact start and stop time of the overall change. The outlier criterion during the initial sweep was changes of 120 knots or more occurring within a period of 60-seconds. In the second sweep, the outlier criterion was changes of 60 knots or more occurring in a single 12-second radar update.

Results

Comparison of Clearances With Computer-detected Changes

Bivariate correlations of tabulated clearances and changes were not an effective means of evaluating concordance because of interval processing (i.e., changes occurring in the interval following the issued clearance). Therefore, clearances were manually paired with their corresponding changes. Several criteria were used for pairing, including temporal proximity of the clearance to the change, the direction of the change, and whether the final recorded altitude, heading, or speed was comparable with the clearance issued.

Altitude. The proportion altitude clearances paired and unpaired with computer-detected changes is shown in Table 2. A total of 71 (84%) of the altitude clearances were paired with changes occurring within the traffic sample processed.

Table 2. *Summary of Altitude Clearance/Change Pairs*

	Altitude Clearances
Paired	71 (84%)
Unpaired	14 (16%)
Total	85

Unpaired altitude clearances. The majority of unpaired altitude clearances were the result of transfer of control. In most cases, the controller issued a clearance for the aircraft to climb or descend to the vertical boundary of the adjacent sector and then immediately transferred control of the aircraft. The change was not detected because the aircraft was no longer under the sector's control when it complied with the altitude clearance (note that all of these changes *were* detected when data from the adjacent sectors were processed, raising the proportion of paired altitude clearances and changes to approximately 97%). Other causes for failure to pair altitude clearances with computer-detected changes included one garbled aircraft identifier, one case of non-compliance (i.e., the pilot never followed the controller's instructions), and one clearance to "stay at present altitude."

Heading. The proportion of heading clearances paired and unpaired with computer-detected changes is shown in Table 3. It is important to note that many of the heading clearances (72%) were paired with corresponding heading changes.

Table 3. *Summary of Heading Clearance/Change Pairs*

	Heading Clearances	
Paired	21	(72%)
Unpaired	8	(28%)
Total	29	

Unpaired heading clearances. Two clearances were issued just before control of the aircraft was transferred to another sector. One could not be paired because the aircraft could not be identified (i.e., the stated aircraft identifier did not correspond with any of the controlled aircraft in or around the sector). In another case, the issued clearance was less than the minimum criterion of 10°. The remaining unpaired clearances were the result of changes that occurred too slowly to be detected.

Speed. The proportion of speed clearances paired and unpaired with computer-detected changes is shown in Table 4. Notice that only slightly more than half (55%) of the speed clearances were paired with speed changes.

Table 4. *Summary of Speed Clearance/Change Pairs*

	Speed Clearances	
Paired	12	(55%)
Unpaired	10	(45%)
Total	22	

Unpaired speed clearances. Most of the unpaired speed clearances were caused by the relationship between speed and other types of changes. Consider the flight data from one of the ZKC samples shown in Figure 1. Just prior to the first data point in the graph, the pilot was given a clearance to reduce speed to 250 knots; by 13:34:26 the aircraft had begun to gradually slow. However, when the aircraft made a slight heading change there was a drastic change in recorded ground speed (13:35:20). As soon as the turn ended, the aircraft's recorded ground speed suddenly dropped to a level suggesting the aircraft had actually continued to slow gradually during the heading change (13:35:32). When the aircraft made another subtle heading shift there was another dramatic increase in recorded ground speed (13:35:44), followed by a sharp decrease in recorded ground speed the instant the turn was completed (13:35:56). Needless to say, the speed change in the example was undetected (due to interference and outlier effects) and unpaired with its corresponding clearance. Analogous changes in recorded ground speeds were found with respect to altitude changes. It was clear from this and numerous similar examples, that recorded ground speed was extremely erratic and unreliable when altitude and heading changes were being made.

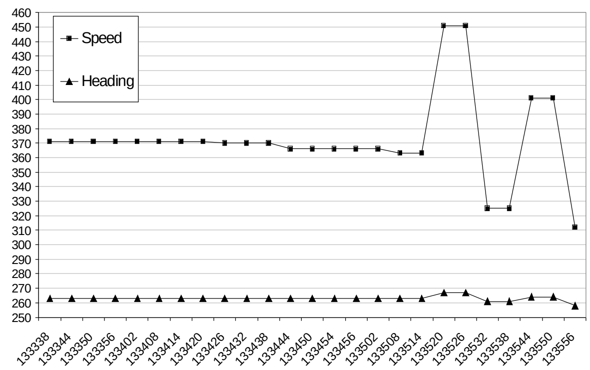


Figure 1. *Example of Changes in Ground Speed Relative to Heading Changes*

Relationship of Clearances and Changes With Subjective Measures of Workload

Multiple regression analysis was employed to examine the relationship between controller clearances and computer-detected changes with subjective controller workload. Two comparative analyses were conducted to examine the amount of variance explained by each set of predictors. A third analysis, using the combined variable sets, examined the amount of shared variance among the predictors. Note that the purpose of these analyses was to evaluate the relationship of the selected variables with the criterion. These variables were not intended to represent a comprehensive list of all possible predictors of subjective controller workload.

As shown in Table 5, the multiple regression model of altitude, heading, and speed clearances produced a multiple R=.59 and accounted for approximately 35% of the variability in ATWIT scores. Both altitude clearances and speed clearances contributed significantly to the model, but heading clearances did not. "In semipartial correlation, the contribution of other IVs is taken out of only the IV. Thus, the squared semipartial correlation expresses the unique contribution of the IV to the total variance of the DV" (Tabachnick & Fidell, 1989, p. 151). The difference between R^2 and the sum of sr^2 for all predictors in the variable set represents shared variance. Therefore, 31% of the variance explained by this variable set was unique, whereas only 4% was shared.

The regression model based on computer-detected changes, shown in Table 6, also produced a multiple R=.59, and accounted for approximately 35% of the variability in ATWIT scores. However, only altitude changes contributed significantly to this model. In this model, 16% of the explained variance was unique and 19% was shared. This indicates that changes were more correlated with one another than were clearances — not surprising given the previously mentioned relationship between changes in ground speed with changes in altitude and heading.

The regression models shown in Tables 5 and 6 demonstrate that both variables sets (i.e., controller clearances and computer-detected changes) are able to explain approximately the same amount of variance in subjective workload. However, this does not necessarily mean that they describe the *same* variance. Squared semipartial correlations of the standard multiple regression analysis of both clearances and changes on ATWIT scores (shown in Table 7) indicate that approximately half (22%) of the 44% explained by the model is shared. (Note that speed changes were excluded from this analysis due to concerns about the accuracy of the variable. It is possible that a larger portion of the explained variance would have been shared had speed changes been included.) Both altitude changes and clearances contributed significantly, but altitude changes explained slightly more unique variance (9%) than did altitude clearances (7%).

Table 5. *Standard Multiple Regression: Clearances on ATWIT Measures (N = 40)*

Model Summary	R	R^2	Adj. R^2	*S.E.*	*F*
	.59	.35	.30	.496	6.49**
Coefficients		sr^2	*b*	*β*	*t*
Altitude		.21	.18	.473	3.45**
Heading		.01	.06	.082	0.59
Speed		.09	.27	.308	2.28*

** p < .01; * p < .05

Table 6. *Standard Multiple Regression: Changes on ATWIT Measures (N = 40)*

Model Summary	R	R^2	Adj. R^2	*S.E.*	*F*
	.59	.35	.29	.497	6.40**
Coefficients		sr^2	*b*	*β*	*t*
Altitude		.12	.15	.474	2.58*
Heading		.02	-.08	-.238	-1.06
Speed		.02	.08	.300	1.11

** *p* < .01; * *p* < .05

Table 7. *Standard Multiple Regression: Clearances and Changes on ATWIT Measures (N = 40)*

Model Summary	R	R^2	Adj. R^2	*S.E.*	*F*
	.67	.44	.36	.472	5.43**
Coefficients		sr^2	*b*	*β*	*t*
Altitude Clearances		.07	.12	.311	2.10*
Altitude Changes		.09	.13	.390	2.38*
Heading Clearances		.01	.07	.093	.67
Heading Changes		.01	-.04	-.104	-.69
Speed Clearances		.04	.19	.218	1.61

** p < .01; * p < .05

Conclusions

The Good. The results of both the tests for concordance and multiple regression analyses demonstrated that altitude clearances and computer-derived altitude changes were strongly related. Though altitude clearances and computer-detected altitude changes did not describe the exact same variance, they were sufficiently related to reduce the amount of unique variance each was able to describe when used in combination in a multiple regression analysis. These results indicate that computer-detected altitude changes might be a viable substitute for altitude clearances in predicting subjective workload.

The Not-So-Bad. The number of heading changes that occurred too gradually to be detected suggested that the heading change algorithms require some revision. A more detailed analysis of heading changes inherent to flight plans and similar sources (i.e., changes in the absence of a clearance) must be conducted before it will be possible to fully determine the accuracy of the algorithms. Although heading changes (and clearances) failed to explain a significant amount of the variance in subjective controller workload, this may not be the case in all traffic samples. Certainly additional analyses using other traffic samples will be necessary to fully evaluate the potential of (revised and improved) computer-detected heading changes as a possible alternative for heading clearances.

The Ugly. On the other hand, ground speeds recorded by the Host computer (and displayed on the controllers' radarscope) proved to be too erratic and unreliable to provide a valid measure of speed changes. Computer-detected measures based on this information cannot be recommended as an acceptable alternative for speed clearances. This is unfortunate, because the results of the regression analysis indicated that controller speed clearances were able to describe a significant amount of unique variance in subjective controller workload. Therefore, it may be worth the time and effort involved to investigate other sources of speed information from which to develop computer-detected measures of speed changes.

References

Bruce, D.S. (1993). An explanatory model for influences of air traffic control task parameters on controller work pressure. *Proceedings of the 37th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 106-112). Santa Monica, CA: Human Factors and Ergonomics Society.

Cardosi, K. (1993). Time required for transmission of time-critical air traffic control messages in an en route environment. *International Journal of Aviation Psychology, 7*, 171-182.

Corker, K.M., Gore, B.F., Fleming, K., & Lane, J. (2000, June). Free flight and the context of control: Experiments and modeling to determine the impact of distributed air-ground air traffic management on safety and procedures. *Proceedings of 3rd USA/Europe Air Traffic Management R&D Seminar.* Napoli, Italy.

Federal Aviation Administration (1995). *Private pilot – airplane practical test standards.* (FAA-S-8081-14). Washington, DC: Federal Aviation Administration Flight Standards Service.

Manning, C.A., Mills, S.H., Fox, C., Pfleiderer, E.M., & Mogilka, H. (2002, May). *Using air traffic control taskload measures and communication events to predict subjective workload.* (Report No. DOT/FAA/AM-02/4). Washington, DC: Office of Aerospace Medicine.

Mills, S.H., Pfleiderer, E.M., & Manning, C.A. (2002). POWER: *Objective activity and taskload assessment in en route air traffic control.* (Report No. DOT/FAA/AM-02/2). Washington, DC: Office of Aerospace Medicine.

Morrow, D., & Rodvold, M. (1998). Communication issues in air traffic control. In M.W.Smolensky & E.S. Stein (Eds.) *Human Factors in Air Traffic Control* (pp.421-456). San Diego, CA: Academic Press.

Porterfield, D.H. (1997). Evaluating controller communication time as a measure of workload. *International Journal of Aviation Psychology, 7,* 171-182.

Rogers, M.D., & Duke, D.A. (1993). SATORI: Situation assessment through recreation of incidents. *The Journal of Air Traffic Control, 35*(4), 10-14.

Stein, E.S. (1985). *Air traffic controller workload: An examination workload probe.* (Report No. DOT/FAA/CT-TN84/24). Atlantic City, NJ: WJH Technical Center.

Tabachnik, B. G., & Fidell, L. S. (1989). *Using multivariate statistics* (2nd ed.). New York: HarperCollins.

A HUMAN FACTORS APPROACH FOR THE ANALYSIS AND THE ENCODING OF AVIATION ACCIDENTS AND INCIDENTS: A VALIDATION STUDY

Pouliquen (Y.) ; **Ferrante (O.)** ; **Jouniaux (P.)** ; **Nicolas (G.)**
Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (BEA), Le Bourget, France

Cabon (P.) ; **Rome (F.)** ; **Wolff (M.)** ; **Mollard (R.)**
Laboratoire d'Anthropologie Appliquée (LAA) - Unité Ergonomie (EA1753) - Université Paris 5, France

Sharing safety information is a key issue to improve aviation safety. Therefore, it appears necessary to have a common way to describe aviation accidents/incidents in order to get consistent data that will be used to produce relevant safety indicators. This implies to use the same taxonomy, the same compatible software to facilitate data sharing, and, more important, a common method to encode occurrences into safety data. The way human factors are taken into account in the database must be improved since statistics usually provided, deal with accident/incident categories and not with their various causes (most of them are human factors related). The BEA in cooperation with the LAA has developed a methodology for the encoding and the analysis of aviation accidents and incidents. This tool has been successfully used during several investigations but still needs to be scientifically validated. This paper aims at putting safety analysis into perspective. It also discusses the methodology that incorporates the Human Factors SHELL model and a validation study.

Introduction

The need for a common and standardized or scientific approach has been highlighted for accident report analysis (Zotov, 2000) and for encoding data from a human factor taxonomy (Casetta et al, 1998). More guidance for reporting has been recently published by ICAO (ICAO, 2003) in addition to what exists in Annex 13. Whereas the facts to be collected are precisely detailed in Annex 13, its appendix only mentions for the analysis: "*Analyze, as appropriate, only the information documented in 1. Factual information and which is relevant to the determination of conclusions and causes*" (ICAO, 2001).

There are several approaches to analyze accidents and incidents. The investigators of the A320 accident of Bahrain (Government of Bahrain, 2002) used a methodology based on the Reason model (Lee and Mulcair, 2003). The Reason model (Reason, 1990) is also used by the US Navy through the Human Factors Analysis Classification System (HFACS) taxonomy to encode occurrences to study error trends across the years to prevent accidents (Shappell and Wiegmann, 2004). A need to validate the results of the encoding process was also taken into consideration (Wiegmann and Shappell, 2001).

The French accident investigation Bureau (BEA), in collaboration with the LAA, has developed an encoding method for occurrence (accident or incident) analysis (Ferrante et al, 2004). This method, which uses the SHELL[1] model (Hawkins, 1987), aims at collecting in an efficient way safety information highlighted during the investigation process and at guiding the investigator into the analysis of the occurrence. The goal is then to be able to disseminate this information through data exchange, safety studies or statistics, mainly focused on human factors and to detect accident precursors. After the development of the method it has been decided to validate it. It consists of verifying the hypothesis that the use of this method harmonizes the determination of causes among investigators and, therefore, increases the reliability of the results that are stored in the database.

This paper summarizes the ADREP causal model structure, the questions raised during an investigation and their associated levels of analysis. It then reviews the methodology stemming from that model and discusses the first results of its validation.

ADREP Causal Model and Associated Levels of Analyses

ICAO adopted the breakdown of an occurrence into a sequence of events which are then described and further explained (see figure 1). This breakdown is useful to classify the different questions that are raised during an investigation and to illustrate the various levels of analysis (Ferrante et al., 2004).

[1] The SHELL model describes a system as the interaction of humans with four elements: Software, Hardware, Environment and Liveware. Each element of the model includes a list of items based on a tree description.

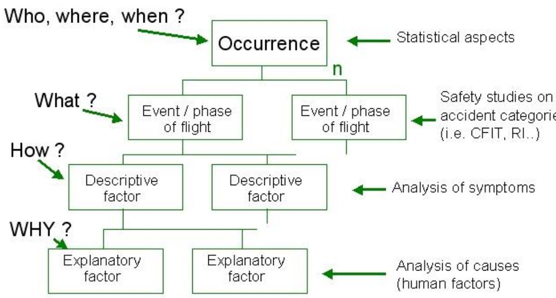


Figure 1. *Four levels of analysis based on ICAO's breakdown of an occurrence*

The first level of analysis covers statistical aspects for different criteria and safety indicators. The first elements gathered right after the notification of an occurrence generally relate to the fields of the flight plan (departure, destination, type of aircraft, date and time). The following questions first asked (who, where, when?) allow to build safety indicators, generally in relation to aircraft, third party damage or injuries. They can be for instance the trend of fatalities in General Aviation or the number of accidents per geographical area. Statistics related to aviation safety are thus mainly based on this type of data, which are only validated during the course of investigations.

The causal approach breaks down an occurrence into a chain of events. Each event is linked to a phase of flight. The number "n" of events depends on the complexity of the occurrence.

The majority of current safety studies are based on these families or categories of events (events correspond to the question "What?" or "which type of occurrence?"). For example, the BEA issued safety studies on fuel starvation events or mid-air collisions (available on www.bea.aero), which correspond to event categories. ICAO and other organizations carried out safety studies on the category of Controlled-flight-into- terrain (CFIT) accidents in the last few years (Flight Safety Foundation, 1996).

ICAO further refines each event by using descriptive factors. These factors mainly refer to aircraft systems, operational or environmental aspects of each event. They correspond to the question "How?". The associated analyses are thereafter based on these identified symptoms. They allow a first level of mitigation measures generally geared to set up palliative actions.

Each descriptive factor is in turn associated with explanatory factors which, as the name indicates, correspond to the question "Why?". These factors reflect the causes of the occurrence. They primarily relate to human factors. These explanatory factors are classified according to the SHELL model which aims at representing the interactions within the aeronautical system. The BEA safety study on the "get-home-itis" factor is an example of an analysis having as a starting point an explanatory factor pertaining to the SHELL model included in the ADREP 2000 taxonomy.

Use of ADREP and ECCAIRS

This latest taxonomy with its 552 explanatory factors represents the outcome of fifty years of investigations throughout the world (Menzel, 2002). It is the third taxonomy version after ADREP 76 (88 factors), ADREP 87 (142 explanatory factors). This material is helpful in tackling systemic issues during an investigation. The clear separation between events and causes, and the fact of having old causes compiled into a taxonomy, help analytical discussions within a team of investigators (national or international). The likelihood of discovering brand new causes is very remote and the ADREP 2000 taxonomy is a natural tool for exploration since it contains an organized collection of all identified events and factors that have, at one time, led to an accident.

The European Commission decided to implement the ADREP taxonomy into a software, ECCAIRS (European Co-ordination Centre for Aviation Incident Reporting Systems) (Cacciabue, 2000). The latest version (ECCAIRS release 4) incorporates ADREP 2000 (and subsequently the SHELL model). Its objective is to facilitate data exchange for analyses on a higher number of occurrences.

However, in addition to a common taxonomy and a common software (ECCAIRS), it is fundamental to have consistent data to prevent biased analyses. This highlights the need of a common methodology to harmonize safety data. Encoding should reflect the report analysis where descriptive and explanatory factors are discussed to elaborate the conclusions.

Two types of practice are currently undertaken to encode an occurrence into ECCAIRS. The first one is done on achieved investigations based on the analysis and the findings of the published reports. This work is generally difficult because an encoder tends to interpret what the investigator had in mind when he wrote his report. It is recommended to stay as objective as possible in order to avoid entering subjective (biased) data in case of interpretation. This approach alters data quality because it is not the person who best knows the case that encodes it. The second one, more appropriate, builds the codification as part of the analysis process to help investigators to elaborate the occurrence causal chain based on factual information and to tackle human factor issues. It has been successfully used during several investigations.

For example, an accident report to a Boeing B737-200 at Tamanrasset (Government of Algeria, 2004) and a serious incident report to a MD83 at Nantes (BEA, 2004), were based on the encoding method. The analyses of these occurrences were undertaken in parallel with encoding and highlighted human factors and systemic issues. In the case of the Tamanrasset accident, this methodology provided tangible material for supporting teamwork, within an international team with people of different backgrounds. It greatly helped putting together the different pieces of the puzzle in the analysis. The main advantage stemmed from the visual tree description of events and factors that illustrate the depth of the investigation. It was thus a powerful and convincing incentive to tackle root causes and their underlying systemic factors.

Principles of the Encoding Method

The main steps of the method are presented in Figure 2 (Ferrante et al, 2004).

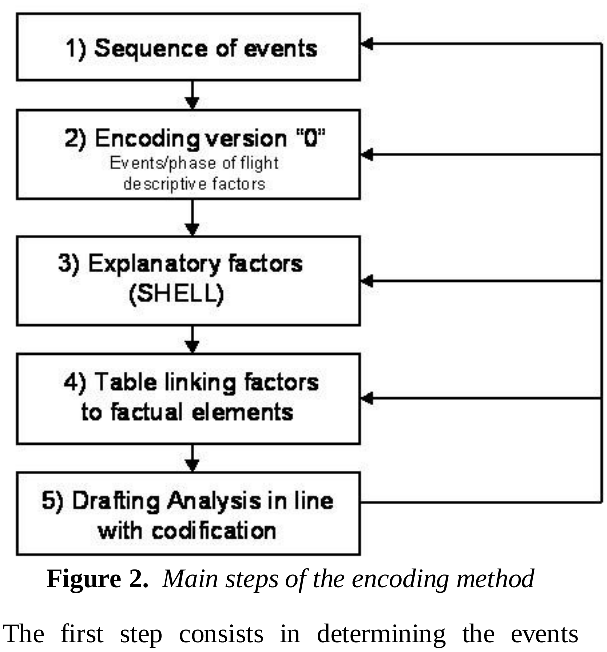


Figure 2. *Main steps of the encoding method*

The first step consists in determining the events leading to the accident/incident. Flight Data Recorder (FDR) and Cockpit Voice Recorder (CVR) data, radar tracks, witness statements and all other information available during the investigation process contain key elements (action, omission, decision, failures, etc.) which will be used to elaborate the sequence of events. Each event is then associated with a phase of flight. A collective approach allows reducing the loss of information and helps dealing with subjective elements, like a witness statement that can be conflicting with factual information.

The descriptive factors precise each event and describe the technical facts and the decisions made by actors which might be later considered as symptoms. One (or more) modifier qualifies each descriptive factor.

The explanatory factors, as they represent the human factor aspects, are chosen within the list given by ADREP 2000, based on the SHELL model. The tree-lists are used as checklists and the explanatory factors are determined after a systematic check. The creation of a table linking the factors to factual elements proved to be very helpful for the justification of the final codification and subsequently the writing of the analysis.

Validation Method for Data Consistency

To validate the encoding method, it is necessary to ensure that it is applied the same way by different investigators.

The validation purpose is to assess the variability of the encoding. Expecting zero variability seems unrealistic. Nevertheless, two investigators using the same method and the same tools should produce similar encoding. The study of the variability of this encoding process should lead to identify the reasons why variability exists. Afterwards, it should be possible to adjust the encoding method by adding enhanced "rules of encoding" to keep the variability as low as possible.

Assessing Variability

To assess encoding variability, the following protocol was applied:

- production of several codifications (sets) per occurrence,
- comparison of the different codifications related to the same occurrence.

A higher number of codifications per occurrence and more occurrences make it easier to bring to light the origin of variability. Therefore it was decided to start with general aviation occurrences, since they are less complex than public transport occurrences and consequently easier to encode in high numbers.

Obviously, this protocol will have to be considered, in a second step, on public transport occurrences, since ADREP 2000 is more dedicated to commercial aviation. The first step, however, consisted of defining criteria to compare different codifications, thus producing initial results.

Production of Several Codifications

During the analysis and encoding steps, two processes can create variability :

- investigators may diverge on the analysis of the same factual information, or
- they may draw the same analysis (same scenario and causes) but without selecting the same elements from ADREP 2000 taxonomy to encode it.

It was first decided to assess the second type of variability, meaning to assess the use of ADREP 2000 by investigators more as an encoding tool than an analytical tool.

Consequently ten occurrences extracted from the General Aviation Bulletin (factual information, analysis and causal factors already available) published by BEA, were given to three separate groups, each composed of one investigator and one human factor specialist. Each group encoded separately these occurrences and highlighted the textual information contained in the report justifying their choices. Doing so, the three sub-levels of codification were covered: events, descriptive and explanatory factors. Then, a comparison of the three resulting codifications was performed in order to quantify and qualify differences.

Comparison of Different Codifications

The next step was to compare the three codifications produced for the same occurrence.

For each occurrence, each pair explained to the others the rationale of encoding the occurrence. During the debriefing, the three groups agreed on a final codification. A significant finding is that the collective approach for encoding helps, as expected, to reduce variability between individuals' interpretation and to produce an agreed final codification.

The following example represents three different codifications and the final one for an accident to a Diamond DA-40 that encountered a power loss during its initial climb. The pilot made a forced landing. The BEA established that the cause of the accident was due to inadequate design of the fuel system. This occurrence was followed by a service bulletin and an airworthiness directive (BEA, 2003). These findings were encoded as illustrated hereafter:

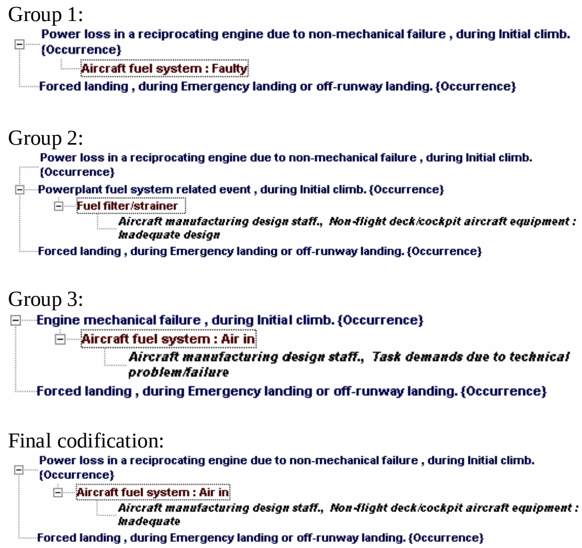


Figure 3. *Comparison and integration of three codifications*

For a given occurrence like the Diamond DA-40 case, all the ADREP 2000 items selected by any of the three groups were listed. For each item, the agreement was scored as follows:

- if selected only by one group, then a "no agreement" was considered,
- if selected by two groups, then a "partial agreement" was considered,
- if selected by the three groups, then a "total agreement" was considered.

Figure 4 shows the results of the agreement between the three groups broken down into the three encoding levels: events, descriptive and explanatory factors.

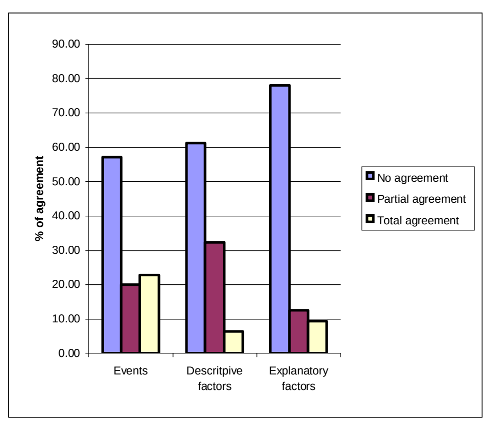


Figure 4. *Percentage of the encoding agreement for three groups*

This comparison shows that variability is higher for explanatory factors than for events and descriptive factors. The nature of the report itself could be a limitation to this validation study since based on a limited analytical narrative.

In addition, this comparison method does not take into account:

- the tree description of ADREP 2000 that leads to score a difference if the items are not strictly identical, although they may belong to the same branch (see figure 5). It would be worth assessing this “proximity” and taking it into account in a further comparison; and
- that a single explanatory factor can be present in two different codifications but without being linked to the same descriptive factor and event. For example, the item “fatigue” can be related to different factors and events.

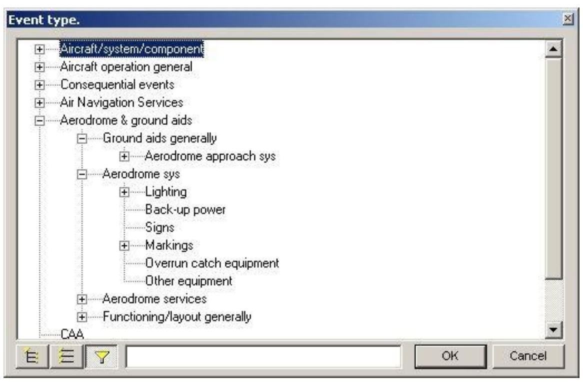


Figure 5. *Tree-list of ADREP events*

This disembodiment of human factor data (Decker, 2001) has to be integrated in an improved comparison process.

These initial results highlight that there are several ways to study differences between codifications. Significant and acceptable differences must be defined. The next step would be to generate a more suitable comparison process. This is still being undertaken.

Initial Explanations and Supplementary Results

The list of points or questions that follows gives initial explanations for this variability, which is related to the use of ADREP 2000 taxonomy through ECCAIRS:

- Investigators do not always check the definition of the ADREP term they select. Therefore, these shortcuts, related to sometimes ambiguous terms, can lead to different interpretations. The on-going learning process has an additional impact on variability.
- A given fact can sometimes be encoded as an event or a descriptive factor.
- Should the breakdown of an occurrence into a chain of events highlight the chronological order of the events or the causal link between events ? This question was answered by placing the causal link as early as possible in the sequence of events, in line with prevention strategies that aim at detecting as early as possible any precursors before they lead to an accident.
- The events and factors section of ADREP 2000 is made of 493 events, 1550 descriptive factors and 552 explanatory factors. Although these numerous elements allow to precisely encode any occurrence, it is sometimes difficult to choose the term that suits the best. Moreover, all investigators do not have the same knowledge of this extended taxonomy.
- When the report is precise enough about a given human factor (e.g. get home-itis, channelized attention, fatigue/alertness), there is generally no variability. When the report does not formally identify a human factor but hints at it, the variability increases because investigators tend to interpret it.

Many of these points are related to training on the use of the method and knowledge of the ADREP 2000 taxonomy. The on-going validation study allows to streamline the methodology and obtain more consistent data.

Conclusion

This pre-validation study has covered a limited number of occurrences from the General Aviation Bulletin where the results of investigations are given in a concise way. On these rather simple cases, a validation protocol was developed. This approach, initially limited to published reports, needs to be enlarged to the direct analysis of factual information, as foreseen for the production of codifications. This represents a time-consuming task for the various groups. It will be even more cumbersome on more complex investigations (with a higher number of events), which generally involve public transport aircraft. This on-going validation study already brought supplementary results to fine-tune the encoding methodological process.

The encoding methodology showed its usefulness on several cases, where a consensus was found for the final codification and for the report analysis. The step by step/iterative approach greatly contributes to its practical use as a tangible support for teamwork. It gives a clear visual understanding of the accident sequence and the associated causes. Investigators have a different knowledge of the extended ADREP 2000 taxonomy. It introduces variability in some codifications and highlights the need for training on the events, descriptive and human factors to share a common understanding of the ADREP definitions.

In the long run, if everybody shares the same concepts, definitions, tools and methods, future prevention measures could be based on standardized and validated results from different countries.

References

Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (2003). Accident on 22 June 2003 at Bernières-d'Ailly (14) to the DA-40 registered F-GNJV. From http://www.bea.aero/docspa/2003/f-jv030622/pdf/f-jv030622.pdf

Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (2004). Serious incident at Nantes on 21 March 2004 to the MD-83 registered SU-BMF operated by Luxor Air. From http://www.bea.aero/docspa/2004/su-f040321a/pdf/su-f040321a.pdf

Cacciabue, P.C. (2000). Human Factors Insight and Reliability Data from Accident Reports: The case of ADREP-2000 for Aviation Safety Assessment. Proceedings of PSAM 5 - International Conference on Probabilistic Safety Assessment and Management. Osaka Nov. 27th - Dec. 1st, 2000.

Casetta, O.P. , Surace, G. , Post, W. (1998). Human Factors data and Accident Reporting Systems for Aviation Safety. Joint Research Centre of the European Commission (JRC) – Institute for the Protection and Security of the Citizen (IPSC) – Technological and Economic Risk Management (TERM) – European Coordination Centre for Aviation Incident Reporting Systems (ECCAIRS)-web site: http://eccairs-www.jrc.it

Decker, S. (2001). The disembodiment of data in the analysis of human factors accidents. Human Factors and Aerospace Safety 1(1), 39-57.

Ferrante, O., Jouniaux, P., Nicolas, G., Loo, T., Nicolas, G., Cabon, P., Mollard, R. (2004). Application of ADREP 2000 taxonomy for the analysis and the encoding of aviation accidents and incidents: a human factors approach. Human Factors and Aerospace Safety 4 (1), 19-48.

Ferrante, O., Jouniaux, P., Loo, T., Nicolas, G. , Cabon, P., Mollard, R. (2004). Application of ADREP 2000 and proposal of a human factors method for the analysis and the encoding of accidents and incidents in aviation. 7th Sino-French symposium Taïpei 1st April 2004.

Flight Safety Foundation (1996). Killers in Aviation : FSF Task Force Presents Facts about Approach-and-landing and Controlled-flight-into-terrain Accidents, Special FSF report, November - December1996.

Government of Algeria (2004). Ministry of Transport, National Commission of Inquiry, Accident on 6 March 2004 at Tamanrasset to the Boeing 737-200 registered 7T-VEZ operated by Air Algérie. From http://www.bea.aero/docspa/2003/7t-z030306a/pdf/7t-z030306a.pdf

Government of Bahrain (2002). Civil Aviation Affairs. Accident investigation report. Gulf Air Flight GF-072 Airbus A320-212, REG. A40-EK on 23 August 2000 at Bahrain. From http://www.bahrainairport.com/arabic/caa/civil_gf072.htm

Hawkins, F.H. (1987). Human Factors in Flight. Aldershot, U.K.: Gower Technical Press.

Lee, R., Mulcair, J. (2003). Can courtesy kill ?. Flight Safety Magazine Australia, 7(5), 26-33. [Electronic version]. From http://www.casa.gov.au/avreg/fsa/03nov/26-33.pdf

ICAO (2001). Annex 13 to the Convention on International Civil Aviation – Aircraft Accident and Incident Investigation - Appendix Format of the final report Accident investigation.- Ninth edition.- International Civil Aviation Organization, Montréal, Canada.

ICAO (2003). Manual of Aircraft Accident and Incident Investigation Part IV Reporting. DOC 9756 AN/965 First Edition. – International Civil Aviation Organization, Montréal, Canada, 2003.

Menzel, R. (2004). ICAO safety database strengthened by introduction of new software, ICAO journal, 59, 3, 2004, 19-26.

Reason, J. (1990). Human Error. Cambridge : Cambridge University Press.

Shappell, S.A., Wiegmann, D.A. HFACS Analysis of Military and Civilian Aviation Accidents: A North American Comparison. Proceedings of the International Society of Aviation Safety Investigators (ISASI) conference, Gold Coast, Australia, 2004.

Wiegmann, D.A., Shappell, S.A. (2001). Applying the Human Factors Analysis and Classification System (HFACS) to the analysis of commercial aviation accident data.- Proceedings of the 11th International Symposium on Aviation Psychology, 2001, 6 p.

Zotov, Dimitri (2000) Scientific methods for accident investigation. Proceedings of the 2000 ANZSASI regional seminar. From http://www.asasi.org/papers/scientific_methods.pdf

RISK FACTORS FOR FATAL GENERAL AVIATION ACCIDENTS IN DEGRADED VISUAL CONDITIONS

Jana M. Price
Loren S. Groff
National Transportation Safety Board
Washington, D.C.

The prevalence of weather-related general aviation (GA) accidents has declined over the past two decades, yet the fatality rate of these accidents remains high. The goal of this study was to examine predictors of fatality within a set of weather-related GA accidents to determine if there are particular factors that contribute to excessively high fatality rates. 3,206 weather-related GA accidents from the National Transportation Safety Board (NTSB) Aviation Accident Database were analyzed using univariate chi-squares and binary logistic regression. A variety of pilot, aircraft, flight, and accident-related factors were evaluated to determine if they increased the odds of pilot fatality. Results suggest that the predictors of fatality in weather-related accidents are similar to those in the greater GA population; but that these factors are more prevalent within weather-related accidents.

Introduction

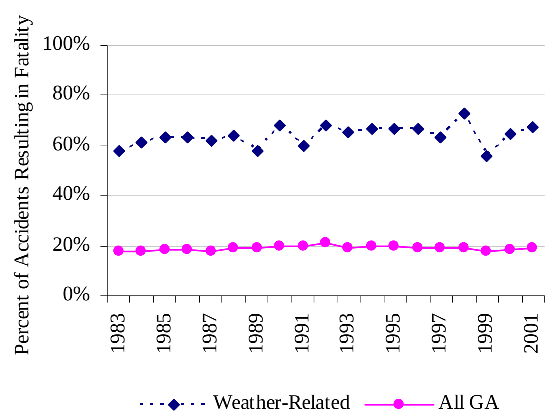
Adverse weather is an ongoing problem for safety in General Aviation (GA) operations, and one that has concerned the National Transportation Safety Board (NTSB) for many years. More than three decades ago, the NTSB published a study of GA accidents suggesting that certain types of weather including low ceiling, rain, and fog were particularly prevalent in fatal accidents involving weather (NTSB, 1968). A separate study found that weather factors such as unfavorable wind, updrafts and downdrafts were associated with nonfatal GA accidents (NTSB, 1976). These findings taken together suggest that degraded visibility is a common factor that separates fatal from nonfatal weather-related accidents.

The link between visibility-related weather and fatal outcomes is paralleled by the fact that accidents that occur in instrument meteorological conditions (IMC) are more likely to result in fatalities than those that occur in visual meteorological conditions (VMC). IMC is defined in the FAA Pilot/Controller Glossary as, "meteorological conditions expressed in terms of visibility, distance from clouds, and ceiling less than the minima specified for visual meteorological conditions." These minima vary by the type of airspace, altitude and light conditions, and are used as thresholds to determine when pilots must operate under instrument flight rules (IFR). While there is a substantial amount of overlap between weather-related accidents and those that occur in IMC, the two are not necessarily coupled; i.e., weather-related accidents may occur during conditions legally defined as VMC.

This study focused specifically on understanding the survivability of weather-related accidents where "weather-related" was operationally defined as those accidents citing visibility-related weather conditions such as fog, rain, snow or low ceilings. The involvement of various weather phenomena in an accident may be determined by reviewing the environmental conditions that were identified by the accident investigator as being a factor in the accident.

A review of NTSB accident data reveals that the proportion of weather-related GA accidents has declined over the years, from more than 11% of all GA accidents in 1983 to less than 6% in 2001. However, as shown in figure 1, the fatality rate for weather-related GA accidents is approximately three times higher than that for all GA accidents, and fatality rates for weather-related GA accidents have been consistently high over the years, ranging between 58% and 72%.

Figure 1. *Proportion of weather-related and all GA accidents that resulted in at least one fatality from 1983–2001.*



Because weather-related accidents are consistently more likely to result in fatality, it is important to search for factors associated with fatal outcomes and, when possible, take steps to mitigate those factors. Studies that have looked at risk factors for fatal injury in airplane accidents across all types of weather have consistently shown that certain variables such as off-airport locations and post-crash fire are strongly associated with higher fatality rates (Li & Baker, 1993; Li & Baker, 1999; O'Hare, Chalmers & Scuffham, 2003). "Off-airport" accidents likely occur when an airplane is at higher altitudes, resulting in crashes with higher speeds and angles of impact. A 1985 report on GA crashworthiness (NTSB, 1985) proposed a "survivable envelope" of impact defined by speeds of 45 knots at 90 degrees of impact angle, 60 knots at 45 degrees, and 75 knots at zero degrees.

Fatalities linked to post crash fire may be caused by burns and smoke inhalation associated with the fire itself, or because high impact crashes result in both fire and death due to deceleration forces. Other factors that have been associated with fatal outcomes in GA accidents include older pilot age, lack of restraint use, and nighttime conditions.

The goal of this study was to examine predictors of fatality within a set of weather-related GA accidents to determine if particular factors contribute to the excessively high fatality rate of those accidents. For example, light conditions may be associated with pilot fatality if the combination of darkness and inclement weather impedes search and rescue operations.

A second possibility is that fatal outcomes in all GA accidents are related to the same basic set of factors, and those factors are more common during flights in degraded visibility. For example, accidents involving multi engine airplanes may have higher fatality rates due to the impact forces associated with higher airspeeds relative to single engine aircraft. If multi-engine airplanes make up a larger proportion of the aircraft flown in inclement weather, it is likely that they would be more represented in weather-related accidents, contributing to their higher fatality rates.

Method

GA airplane accidents citing one or more visibility-related condition, for the period of 1983–2001, were extracted from the NTSB Aviation Accident Database. Independent variables were chosen based on hypothesized relationships between the variable and the likelihood of survival, though the selection of variables was limited to those that were well represented in the database.

Pilot-related independent variables included pilot age, highest certification (student, private, commercial or air transport), instrument rating (yes or no), total flight hours, seatbelt use (yes, no or unknown), and shoulder restraint use (yes, no or unknown). Aircraft-related variables included number of engines (single or multi-engine) and airplane construction (amateur-built or manufactured). Flight and accident related variables included light condition (light or dark), presence of fire (yes or no), presence of explosion (yes or no), collision as the first occurrence (yes or no), and phase of flight (standing/taxi, takeoff, climb, cruise, descent, maneuvering, approach, go-around or landing).

The dependent variable was the case-fatality rate, or the proportion of cases in which the pilot was fatally injured. Univariate chi square tests were used initially to assess the effect of each independent variable on the case fatality rate. Binary logistic regression was then used to evaluate the combined effect of the independent variables and to assess the significance of each predictor in the presence of all others. All data analyses were performed using the SPSS software.

Results

Of the 37,681 GA airplane accidents that occurred between 1983 and 2001, 3,206 or 8.5% were weather-related. Within this group, 71.5% occurred in IMC, 60.2% involved restrictions to visibility such as fog or haze, and 48.5% occurred during precipitation.

Univariate Analyses

The overall case fatality rate for the sample was 62.4%, and case fatality rates for all levels of each independent variable are shown in Tables 1 and 2. All of the pilot-related independent variables produced significant chi-square findings; a private pilot license, no instrument rating, fewer total flight hours, and older pilot age were associated with higher case fatality rates. Non-use of restraints such as seatbelts and shoulder harnesses was also associated with higher case fatality rates; however, due to the large number of cases where restraint use was missing or unknown, restraint use variables were not included in the multivariate analyses.

Neither of the aircraft-related variables (number of engines or aircraft construction) produced significant chi square findings, but all flight and accident-related factors significantly influenced case fatality rates. Darkness, presence of fire, presence of explosion, and collision as first occurrence were associated with higher case fatality rates.

Multivariate Analysis

Binary logistic regression incorporating pilot, aircraft, flight and accident-related variables resulted in a significant omnibus finding supporting the overall model. When controlling for all other variables, the individual pilot-related factors associated with a higher risk of fatality included older pilots and those holding a private pilot license, as compared to other types of pilot certificate. For example, pilots aged 34–43 were 1.38 times more likely to die in an accident than younger pilots, and the odds of fatality increased further for pilots older than 43. Also, pilots whose highest level of certification was the private license (OR = 1.81) were more likely to die compared to the reference group of those with air transport licenses.

Accidents involving multi-engine aircraft were 1.55 times more likely to result in a pilot fatality than those involving single-engine aircraft. Darkness (OR = 1.68), and the presence of fire (OR = 5.87) also significantly increased the odds of pilot fatality. For phase of flight, accidents that occurred during the standing/taxi and landing phases were least likely to result in a fatality. Accidents during the climb and cruise phases had the highest odds of fatality, each being greater than 30 times more likely to result in a pilot fatality than accidents that occurred during the standing/taxi phase.

Table 1. *Case fatality rates, odds ratios, and confidence intervals for pilot-related factors.*

Variable	Number of Pilots Involved	Number of Fatally Injured Pilots	Case Fatality Rate (%)	Odds Ratio	95% CI
Pilot Age					
16-33	651	337	51.8	Ref	--
34-43	706	430	60.9	1.38*	1.07, 1.79
44-53	883	575	65.1	1.90**	1.47, 2.46
>53	939	646	68.8	2.36**	1.79, 3.12
Highest Certification					
Student	98	47	48.0	0.87	0.45, 1.67
Private	1858	1259	67.8	1.81**	1.20, 2.72
Commercial	934	522	55.9	1.08	0.75, 1.56
ATP	293	156	53.2	Ref	--
Instrument Rated					
No	1419	928	65.4	Ref	--
Yes	1773	1068	60.2	1.14	0.90, 1.45
Total Flight Time					
0-247	568	351	61.8	1.42	0.97, 2.08
248-825	887	573	64.6	1.29	0.95, 1.75
826-2799	870	546	62.8	1.11	0.85, 1.46
>2799	771	444	57.6	Ref	--
Seatbelt Used					
No	30	28	93.3		
Yes	2569	1434	55.8		
Unknown	582	517	88.8		
Shoulder Harness Used					
No	781	483	61.8		
Yes	1297	625	48.5		
Unknown	1093	861	78.8		

* p < .05, ** p < .01

Table 2. *Case fatality rates, odds ratios, and confidence intervals for aircraft, flight, and accident-related factors.*

Variable	Number of Pilots Involved	Number of Fatally Injured Pilots	Case Fatality Rate (%)	Odds Ratio	95% CI
Number of Engines					
Single Engine	2536	1571	61.9	Ref	--
Multi Engine	662	425	64.2	1.55**	1.20, 2.00
Amateur-Built					
No	3124	1946	62.3	Ref	--
Yes	76	52	68.4	1.32	0.74, 2.35
Light Condition					
Light	2062	1209	58.6	Ref	--
Dark	1125	777	69.1	1.68**	1.39, 2.02
Presence of Fire					
No	2452	1332	54.3	Ref	--
Yes	726	645	88.8	5.87**	4.28, 8.05
Presence of Explosion					
No	2913	1733	59.5	Ref	--
Yes	242	220	90.9	1.49	0.86, 2.58
Collision as First Occurrence					
No	2515	1533	61.0	Ref	--
Yes	686	465	67.8	1.23	0.98, 1.54
Phase of Flight at First Occurrence					
Standing/Taxi	19	1	5.3	Ref	--
Takeoff	254	146	57.5	13.58*	1.74, 106.21
Climb	203	155	76.4	38.99**	4.93, 308.30
Cruise	1375	946	68.8	31.18**	4.05, 240.25
Descent	124	87	70.2	28.50**	3.56, 228.35
Maneuvering	348	234	67.2	27.49**	3.52, 214.70
Approach	560	323	57.7	12.08*	1.56, 93.53
Go-Around	86	60	69.8	20.86**	2.56, 170.25
Landing	194	8	4.1	0.56	0.06, 4.64

* $p < .05$, ** $p < .01$

Discussion

Over the past two decades, weather-related GA accidents have been consistently more likely to result in fatalities than GA accidents overall. The goal of this research was to determine if particular factors within weather-related accidents are uniquely predictive of pilot fatality.

Similar to previous research (Li & Baker, 1999; O'Hare et al., 2003), some of the most predictive factors were related to the accident occurrence. For example, the climb and cruise phases of flight were associated with the highest odds of pilot fatality, similar to other research findings in which pilot fatality was linked to "off airport" accidents. In both cases, accidents that occurred mid-flight were linked to severe outcomes, presumably due to higher airspeeds. Accidents that occurred at night were also more likely to result in pilot fatality.

The presence of fire increased the risk of pilot fatality by nearly six times. The presence of explosion and collision were associated with higher case fatality rates, but did not significantly increase the odds of pilot fatality when controlling for all other predictors, possibly due to the correlation of these variables with the presence of fire.

Among pilot-related predictors, older pilots and those with a private license had increased odds of fatality. The finding that older pilots were more likely to die in weather-related crashes was consistent with previous research and was presumably due to increasing frailty with age.

Of the two airplane-related factors, only number of engines was significant in the multivariate analysis, with accidents involving multi-engine aircraft approximately 1.6 times more likely to result in pilot fatality. While this finding is difficult to explain without acknowledging pilot and operational differences, the association between aircraft size and pilot fatality was likely a consequence of the fact that multi-engine aircraft fly faster than single-engine aircraft, resulting in higher accident impact forces.

In sum, the majority of factors that were predictive of fatality among weather-related accidents, such as fire, dark conditions, intermediate phases of flight, multi-engine aircraft, and older pilot age, have also been linked to higher fatality rates among all GA accidents. These findings suggest that weather-related accidents are caused by the same underlying set of factors as non-weather accidents, but those factors are more prevalent in weather accident scenarios. Initial analysis of a corresponding set of non-weather accidents seems to support this idea. For example, more than 35% of the weather-related accidents in this study occurred at night, which is more than four times higher than the proportion of nighttime non-weather GA accidents (7.6%) over the same time period. Similarly, multi-engine airplanes made up 20.6% of weather-related accidents, but less than 9% of non-weather accidents.

There were, however, a few risk factors that uniquely predicted pilot fatality in weather-related accidents. Pilot-related factors, such as highest level of certification and flight hours, were specifically related to fatal accident outcomes. Among licensed pilots, private pilots and those with fewer flight hours were more likely to die than air transport pilots and those with more flight hours. These findings were not evident in research by Li and Baker (1999), who found few differences among pilots with private, commercial, and air transport certificates, and higher fatality rates among pilots with the greatest number of flight hours.

The relationship between pilot experience and accident outcomes is likely mediated by accident circumstances. For example, private pilots and/or those with fewer flight hours may be more susceptible to weather-induced problems, such as spatial disorientation or loss of control, which typically result in serious accidents. This finding points to a need to examine in greater detail the relationships between pilot characteristics and specific accident circumstances.

Previous laboratory- and survey-based studies of IMC or weather-related accidents have focused on factors such as pilots' ability to detect and evaluate deteriorating visibility conditions (Weigman, Goh, & O'Hare, 2002), their ability to make decisions regarding inclement weather (Burian, Orasanu & Hitt, 2000), and the availability of weather information during flight (Latorrella & Lane, 2002). Similar to the analysis presented here, these studies focused primarily on identifying commonalities among accident pilots rather than identifying those factors unique to accident involvement. To date, few field studies have investigated the factors that distinguish weather-related accident pilots from those pilots able to operate successfully in similar conditions—an area that would benefit from continued research.

References

Burian, B. K., Orasanu, J., & Hitt, H. (2000). Weather-related decision errors: Differences across flight types, *Proceedings of the IEA 2000/HFES 2000 Congress*, 22-25.

Latorella, K. & Lane, S. (2002). *General aviation pilots' perceived usage and valuation of aviation weather information sources.* NASA/TM-2002-211443.

Li, G. & Baker, S. P. (1993). Crashes of commuter aircraft and air taxis: What determines pilot survival? *Journal of Occupational Medicine, 35*(12), 1244-1249.

Li, G. & Baker, S. P. (1999). Correlates of pilot fatality in general aviation crashes, *Aviation Space and Environmental Medicine, 70*(4), 305-309.

National Transportation Safety Board (1968). *An analysis of U.S. general aviation accidents involving weather as a cause/related factor 1966.* NTSB.

National Transportation Safety Board (1976). *Nonfatal, weather-involved general aviation accidents*, NTSB-AAS-76-3.

National Transportation Safety Board (1985). *General aviation crashworthiness project: Phase two—Impact severity and potential injury prevention in general aviation accidents*, NTSB/SR-85/01.

O'Hare, D., Chalmers, D., & Scuffham, P. (2003). Case-control study of risk factors for fatal and nonfatal injury in crashes of civil aircraft. *Aviation, Space and Environmental Medicine, 74*(10), 1061-1066.

Weigmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules flight into adverse weather, *Human Factors*, *44*(2), 189-197.

FLIGHT TEST EVALUATION OF SITUATION AWARENESS BENEFITS OF INTEGRATED SYNTHETIC VISION SYSTEM TECHNOLOGY FOR COMMERCIAL AIRCRAFT

Lawrence J. Prinzel III, **Lynda J. Kramer**, **Jarvis J. Arthur III**,
Randall E. Bailey, **J. Raymond Comstock**

NASA Langley Research Center
Hampton, VA

Research was conducted onboard a Gulfstream G-V aircraft to evaluate integrated Synthetic Vision System concepts during flight tests over a 6-week period at the Wallops Flight Facility and Reno/Tahoe International Airport. The NASA Synthetic Vision System incorporates database integrity monitoring, runway incursion prevention alerting, surface maps, enhanced vision sensors, and advanced pathway guidance and synthetic terrain presentation. The paper details the goals and objectives of the flight test with a focus on the situation awareness benefits of integrating synthetic vision system enabling technologies for commercial aircraft.

Introduction

A "synthetic vision system" is an electronic means of displaying the pertinent and critical features of the environment external to the aircraft through a computer-generated image of the external scene topography using on-board databases (e.g., terrain, obstacles, cultural features), precise positioning information, and flight display symbologies that may be combined with information derived from a weather-penetrating sensor (e.g., runway edge detection, object detection algorithms) or with actual imagery from enhanced vision sensors.

NASA Synthetic Vision System Project

NASA's Synthetic Vision Systems (SVS) project is developing technologies with practical applications that will eliminate low visibility conditions as a causal factor to civil aircraft accidents while replicating the operational benefits of clear day flight operations, regardless of the actual outside visibility condition. A major thrust of the SVS project involves the development/demonstration of affordable, certifiable display configurations that provide intuitive out-the-window terrain and obstacle information with advanced pathway guidance. The SVS concept being developed at NASA encompasses the integration of tactical and strategic Synthetic Vision Display Concepts (SVDC) with Runway Incursion Prevention System (RIPS) alerting, real-time terrain database integrity monitoring equipment (DIME), and Synthetic Vision Sensors (SV-Sensors), using an enhanced weather radar for real-time object detection, runway confirmation, and database integrity monitoring.

Previous flight tests (Glabb et al., 2003; Kramer et al., 2004) of SVS have primarily focused on the general use and utility of SVS for providing flight critical guidance and improved terrain/situation awareness. The research objectives of these previous flight tests also focused on SVS implementation issues, such as display requirements (e.g., size, content, and format) and on the development of SVS enabling technologies (e.g., RIPS, EVS, and DIME).

While research to date has proven that precision navigation and on-board databases can provide the primary framework for substantial improvements in terrain/situation awareness with SVS, independent integrity monitors are envisioned as an integral component of a Synthetic Vision System to meet flight-critical safety requirements. This functionality is being developed by NASA and others to utilize existing on-board sensors (e.g., weather radars, high quality radar altimeters) to facilitate implementation. Specific on-board integrity functions include independent air-to-air, air-to-ground, ground-to-ground, and ground-to-air traffic and object/obstacle detection and surveillance, a runway incursion monitoring, and database integrity and registration (navigational position confirmation via terrain feature extraction). Additionally, SVS concepts are being developed to augment and complement the independent capabilities of weather-penetrating, enhanced vision imaging sensors during low visibility landing and surface operations conditions. These technologies form the basis for monitoring the dynamic flight environment and thereby supplementing the synthetic world with real-time, direct measurement of the surrounding terrain and air/ground traffic for flight-critical applications.

SVS Integrated Flight Test

A flight test evaluation was jointly conducted (in July and August 2004) by NASA Langley Research Center and Gulfstream Aerospace Corporation under NASA's Aviation Safety and Security (AvSSP), Synthetic Vision System program. A Gulfstream G-V aircraft was flown over a 3-week period in the Reno/Tahoe International Airport (RNO) local area and an additional 3-week period in the Wallops

Flight Facility (WAL) local area to evaluate an integrated Synthetic Vision System concept, including real-time, integrity monitoring functions.

Flight Test Objectives

The primary G-V Synthetic vision Integrated Technology Evaluation (GVSITE) flight test objective was to evaluate the utility and acceptance of an integrated Synthetic Vision *System* intended for commercial and business aircraft in a terrain-challenged operational environment.

The integrated SV system included computer-generated terrain presented on Primary Flight Displays (PFD) and Electronic Attitude and Direction Indicators in place of the conventional blue sky and brown ground; monochrome textured terrain presented on Head-Up Displays (HUD); plan view or perspective views of computer-generated terrain and obstacles on Navigation Displays (ND); and datalink, sensors, and algorithms to provide and verify required information for display. In addition, symbology and algorithms designed as integrity monitors and detection/surveillance monitors to enhance pilot situational awareness during surface and landing phase operations, and prevent or alert to potential runway incursions, was also part of the SV system tested during the GVSITE flight test.

Method

Pilot Participants

Ten evaluation pilots (EPs), representing the airlines, a major transport airport manufacturer, the Federal Aviation Administration, and the Joint Aviation Authority, flew research flights totaling approximately 45 flight test hours. One hundred and forty-five flight test runs were conducted to evaluate the NASA SVS concepts at WAL (8 pilots) and RNO airports (7 pilots). Five of the ten EPs flew at both test locations. All participants were HUD qualified.

Test Aircraft

The flight test was conduced using a Gulfstream G-V aircraft. The left seat of the G-V was occupied by the EP and the right seat was occupied by a Gulfstream Safety Pilot (SP). The left seat included in the installation of two 8"x8" (approximately 768x768 pixel resolution) head-down displays for evaluation of the PFD and ND concepts (Figure 1), an overhead Rockwell-Collins HGS-3300 HUD for evaluation of head-up concepts, and a voice recognition and speech (VRS) system for the pilot-vehicle interface to the SV displays. A vision restriction device (VRD) was placed in the left-seat forward windscreen to block the EP's forward vision and thus simulate Instrument Meteorological Conditions (IMC) when needed experimentally. The VRD was removed no lower than 200 ft. above field elevation.

Figure 1. *GulfStream-V SVS Head-Down Displays*

Runway Incursion Prevention System

Real-time, RIPS algorithms (from NASA/LaRC in-house developments and the Rannoch Corporation) and RIPS display concepts were integrated into the Synthetic Vision Display Concepts for GVSITE. RIPS receives data on potential airborne and surface intruders through datalink and onboard sensors, processes the data through RIPS algorithms and known aircraft position to detect potential hazards, and interfaces through cockpit displays and communication systems to warn the crew. Only the NASA LaRC algorithms results are discussed in the paper.

SV Sensors

A modified WxR-2100 multi-mode weather radar with mounting trays, waveguide with a matched load termination, wiring harness, control head, pedestal, and antenna was installed in the G-V to support SV-Sensor research objectives. During the flight test, the radar operated in one of four modes: (a) weather radar – standard weather radar functionality; (b) runway outline identification – ground clutter returns were analyzed with aircraft navigational state data to provide an estimate runway position; (c) terrain feature extraction - ground clutter returns were provided to the DIME as source data, (d) air-to-ground obstacle detection – radar data processing was used in an attempt to identify objects and obstacles on the active runway while on approach.

Database Integrity Monitoring Equipment

A real-time digital terrain elevation data (DTED) integrity monitoring capability was designed to detect statistically significant differences between sensed terrain data and the stored DTED through two DIME concepts:

1. Using inputs from the ship's standard radar altimeter and an internal GPS Wide Area Augmentation System (WAAS) receiver, an estimate of DTED integrity was generated in real-time. This DIME-provided integrity measure was used to create a loss-of-integrity alert which was part of the Synthetic Vision Display concepts. This integrity alert function was experimentally tested.
2. A forward-looking monitor was also tested that makes use of WxR2100 and inertial reference unit (IRU) measurements to complement the radar altimeter-based integrity monitor.

An experimental GPS bi-static radar equipment was also installed in the DIME rack to collect data to support subsequent algorithm development for a possible third database integrity method.

Enhanced Vision Sensor

Enhanced Vision System (EVS) capability was provided by the standard G-V Kollsman Forward Looking InfraRed (FLIR) camera. The cryogenically-cooled FLIR camera operates in the low-to-mid IR wavelengths using a sensor with approximately 320 Horizontal x 240 Vertical pixel resolution. The EVS generated an RS-170 video signal which was up-converted to an RS-343 video signal for the Flight Dynamics HUD through a Folsom scan converter.

Experimental Display Conditions

Four display conditions (Figure 2) were evaluated while EPs performed approaches and departures at RNO and WAL airports:

1. The first display condition (Baseline) utilized both the head-down and head-up research displays. The head-down displays represented a conventional PFD and ND. The ND was a co-planar display with a map-centered Terrain Awareness and Warning System (TAWS) display and a vertical situation display (VSD). No synthetic terrain information was presented on either the head-up or head-down displays in the Baseline condition.
2. The second display condition (Baseline FLIR) had the same head-down PFD and ND concepts as the Baseline display condition, but it included FLIR on the raster channel of the HUD.
3. The third display condition (Advanced SVS) utilized the head-down displays and the HUD. In addition to the conventional flight symbology typically found on a PFD and HUD, these displays also included advanced pathway guidance and terrain information using a combination of photo-realistic and elevation-based shading texturing. The ND had terrain information in addition to the TAWS warning and caution overlays and VSD. A surface guidance map display was presented on the navigation display for scenarios with surface operations. The surface map showed the ATC taxi route and active runways and provided alerting of non-normal events (e.g., cross hold-line of active runway, off-route)
4. The fourth display condition (Advanced SVS – No HUD) was exactly the same as the Advanced SVS display condition but it did not employ the HUD. Hence, the EPs primary flight reference was solely head-down.

Flight Evaluation Tasks

At each flight test location (WAL, RNO), EPs flew multiple scenarios which included: approach with wave-off to a departure; approach and landing; taxi operations; low-speed rejected take-off; and takeoff and departure. In addition to nominal approach and departure tasks, there were non-normal runs flown with each display condition which included runway incursion (RI) scenarios and database integrity monitoring scenarios. The RI scenarios included potential incursions with either a Beech King Air (Be-200) or a specially-equipped recreational vehicle during approach, surface, and departure operations. These scenarios were pre-briefed and carefully staged to ensure safety of flight and maximize masking of the RI scenario from the EP. The database integrity monitoring scenarios purposefully introduced a SV database offset either laterally or vertically with the real world. The pathway guidance was always correct and the EPs were instructed to fly with respect to the guidance and not the database image. The EPs were instructed to fly each approach as precisely as possible using the display information available to them, as the effect of the display information on the EPs ability to fly the approaches would be quantitatively and qualitatively evaluated. In addition, the EPs were instructed to taxi as close as possible to the centerline of the taxiway, using a ground speed between 15 and 20 knots with a target speed of 18 knots.

Runway Incursion Scenarios

There were seven runway incursions scenarios used for evaluation of RIPS alerting and surface map displays. The scenarios were:

1. Crossing Runway – Departure of test aircraft and departure of incursion aircraft (WAL, RNO)
2. Crossing Runway – Departure of test aircraft and arrival of incursion aircraft (WAL)

3. Crossing Runway – Arrival of test aircraft and departure of incursion aircraft (WAL, RNO)
4. Crossing Runway – Arrival of test aircraft and arrival of incursion aircraft (WAL)
5. Taxi crossing/departure – Taxi across hold line of test aircraft during departure of incursion aircraft on active runway (WAL, RNO)
6. Take-Off Hold/Arrival --- Incursion aircraft on short final and test aircraft at take-off position (WAL)
7. Arrival/Take-Off Hold --- Test aircraft on short final and incursion aircraft at take-off position (WAL, RNO)

Results

Approach Phase, Flight Technical Error

The independent variables were display type (Baseline, Baseline FLIR, Advanced SVS, Advanced SVS-No HUD), path type (Sparks East 16R, Sparks North 16R, South Hills East 34L, and South Hills South 34L), and pilot. The dependent measures were RMS lateral path error and RMS vertical path error. The calculation for RMS path error began on each run when the pilot entered the tunnel the first time. Display type, path type, pilot, and the second order interactions between the main factors were not significant ($p>.05$) for either measure. The pilot performance results are not surprising and are supported by past research (Kramer et al., 2004, Prinzel et al., 2004). Each display concept utilized the same pursuit guidance control laws and symbology (i.e., the flight path marker, integrated single cue guidance symbol and path deviation indicators which commanded the pilot where to fly). The addition of the tunnel concepts in the advanced display formats were not significant in this quantitative path performance data, but did, as shown in the following, influence the subjective workload and SA measures. The FTE results also do not neatly include the influence guidance and tunnel symbology with off-path starting conditions, because it was not possible to precisely control the run-start conditions in the dynamic air traffic/flight test environment; thus, the FTE results were normalized by using the tunnel intercept condition (whether the tunnel was explicitly shown or not) to begin the FTE "scoring."

Approach Phase, Mental Workload

There were no statistically significant differences for the Air Force Revised Workload Estimation Scale amongst the display concepts, ($p > .05$). Pilots rated the workload from "light" (Advanced SVS) to "moderate activity" (Baseline). However, SWORD ratings during approach revealed that pilots rated the baseline condition significantly higher in mental workload than the other three display conditions ($F(3,33) = 8.470$, $p < .05$). The baseline condition is the only display configuration that doesn't explicitly have terrain information on the PFD or HUD.

Approach Phase, Pilot Situation Awareness

The SA-SWORD analysis revealed two unique subsets for display concept comparisons for situation awareness during approach ($F(3,27) = 8.188$, $p < .05$): (1) advanced SVS (highest) and (2) advanced SVS – no HUD, Baseline with FLIR, and Baseline (lowest). The advanced configuration differs from the other three configurations, principally by having terrain information on the PFD and HUD.

Surface Operations, Workload

For surface operations, there were three unique subsets for SWORD ratings ($F(3,30) = 23.196$, $p < .05$): (a) Advanced SVS (lowest), (2) Advanced SVS – no HUD, and (3) Baseline with FLIR and Baseline (highest). Two prominent display configuration differences influence the surface operations results – the presence of the Electronic Moving Map (EMM) in the advanced display concept and surface guidance symbology and the presence of a HUD.

Surface Operations, Situation Awareness

There was also a significant effect found for SA-SWORD for surface operations ($F(3,33) = 14.075$, $p < .05$) revealing three unique subsets for display concept comparisons for situation awareness for surface operations: (1) advanced SVS (highest); (2) Advanced SVS – no HUD and Baseline with FLIR; and (3) Baseline with FLIR and Baseline (lowest). The situation awareness results mirror those of the workload results, signifying the importance of advanced guidance and situation information on a HUD for ground operations. The importance of situation information is further highlighted by pilot subjective reports of improved SA for ground operations using the EMM as highlighted in the following.

Pilots rated their situation awareness very high for surface operations when using the surface map displays, considered an essential part of the integrated NASA synthetic vision system, compared to surface operations using the baseline surface display. Post-experiment questions were asked of pilots regarding surface operations and situation awareness using the surface map display and alerting. For each question, pilots rated 1 (completely disagree) to 7 (completely agree) on a Likert scale in terms of agreement for the following questions (Figure 3):

Q1: Where am I? "The display concept provides sufficient awareness of my ownship position with respect to runways, taxiways, and stationary objects."
Q2: Where am I relative to Other Moving Objects?" The display concept provides sufficient awareness of my ownship position with respect to moving traffic, such as vehicles and other aircraft."
Q3: What is the status of surfaces in the movement area? "The display concept provides sufficient awareness of the status of taxi and runway surfaces."
Q4: Where am I relative to my route/destination? "The display concept provides sufficient awareness of my cleared route."
Q5: What control inputs should I make to maintain my cleared route? "The display concept provides sufficient guidance cues needed to follow my cleared route."

Figure 2 graphically demonstrates that pilots rated the EMM display significantly higher for situation awareness across all five questions that addressed a different facet of SA. On average, pilots completely agreed with the statements that the EMM significantly enhanced awareness of ownship position and those of other aircraft and vehicles, cleared taxi route, and active runways and surface information. Pilot unanimously considered the EMM to be an essential and needed cockpit display that would substantially enhance aviation safety and efficiency.

Runway Incursion Prevention

Pilots encountered seven runway incursion scenarios at WAL and 4 incursion scenarios at RNO. A total of 82 experimental runs were conducted at WAL and 60 runs were conducted at RNO. Overall, the RIPS algorithm results are very promising (data analysis is on-going), showing successful detection and minimal false alarms (Jones, in press).

In terms of the situation awareness provided by RIPS, pilots rated the RIPS alerting to be better than the baseline conditions for "likelihood of detecting and preventing a runway incursion." The inclusion of RIPS alerting was rated 6.96/7.0 (very high likelihood) compared to only 2.64/7.0 (low likelihood) for the baseline conditions. 9/10 pilots reported that the incursion alerts were provided in a timely manner and felt that RIPS significantly enhanced RI safety compared to current technology and procedures (cockpit, ground, ATC). After familiarization, the majority of the pilots (9/10) trusted the alerting and initiated a go-around or evasive action on the ground to avoid a runway incursion. Only one pilot needed to first confirm the hazard before initiating a go-around.

Integrity Monitoring

Pilots were asked to provide two ratings, one on the effectiveness and one on the essentialness, on the presentation of NOTAM alerts (e.g., NOTAM tower, closed rwy) and DIME alerts for a synthetic vision system. Pilots used a Likert rating scale (1-7) to rate the effectivenss and essentialness of the NOTAM and DIME information presentations. An average rating of 4.2 (moderately effective/essential) was reported for NOTAM tower alerts but pilots rated NOTAM closed rwy alert presentation to be completely effective and essential (7.0). For DIME alerts, pilots rated the information presentation as being highly effective (6.42) and completely essential (7.0).

Pilot Preference

Pilots were asked to rank order display concepts in terms of (a) pilot performance and flight path awareness and (b) pilot preference for IMC approaches. A Friedman test ($p < .05$) evinced a significant ranking for both questions in the order of: (1) advanced SVS (highest); (2) Advanced SVS – no HUD; (3) Baseline with FLIR; and (4) Baseline lowest). Pilots also provided a number of useful comments that have been used to guide subsequent and future SVS developments. Overall, however, pilots unanimously applauded the safety and situation awareness benefits of the NASA integrated synthetic vision system.

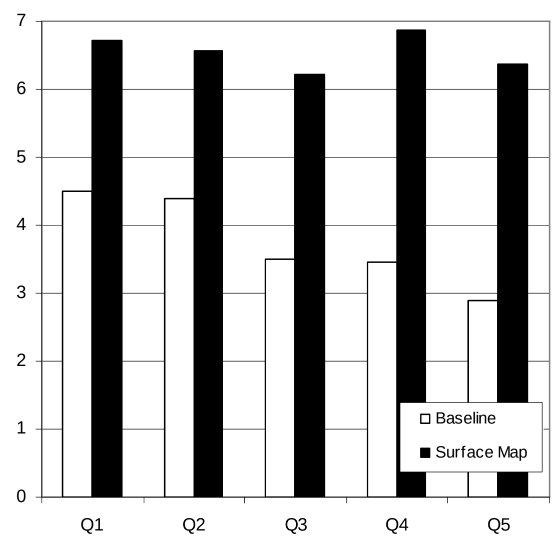


Figure 3. *Situation awareness for surface operations*

Conclusions

The flight test marked the first time NASA's technologies have been integrated as a complete system incorporating synthetic terrain primary flight and navigation displays, advanced weather radar object detection, synthetic vision database integrity monitoring, refined dynamic tunnel and guidance concepts, surface map displays, and the runway incursion prevention system (RIPS). The results showed the efficacy of the NASA Synthetic Vision System to significantly enhance pilot situation awareness for runway traffic and terrain, and substantially better pilot acceptability and trust due to integrated integrity monitors and enhanced vision sensors.

Future Research

The NASA AvSSP SVS project has since conducted an experiment examining the efficacy of 3-D exocentric multi-mode SVS navigation displays with significant positive results. Future research will focus on (1) enhancement of the dynamic tunnel concept to provide 4-D required time of arrival and required navigation performance, (2) crew coordination human factors research using SVS, (3) exocentric dynamic 3-D SVS navigation displays for approach and missed approach rehearsal, (4) military applications of synthetic vision, (5) advanced display media, and (6) integration of SVS with other emerging NASA cockpit information displays.

References

Glabb, L.J., Kramer, L.J., Arthur, T., & Parrish, R.V. (2004). Flight test comparison of synthetic vision display concepts at Dallas/Fort Worth International airport. NASA-TP-2003-212177.

Kramer, L.J., Prinzel, L.J., Bailey, R.E., Arthur, J.J., & Parrish, R.V. (2004). Flight test evaluation of synthetic vision concepts at a terrain challenged airport. NASA-TP-2004-212997.

Jones, D. (in press). Runway incursion prevention system testing at the Wallops flight facility. International Society for Optical Engineering. Orlando, FL: SPIE

Prinzel, L.J., Arthur, J.J., Kramer, L.J., & Bailey, R.E. (2004). Pathway concepts experiment for head-down synthetic vision displays. SPIE. In Proceedings of SPIE, Enhanced and Synthetic Vision 2004, Editor: Jacques G. Verly, Volume 5424.

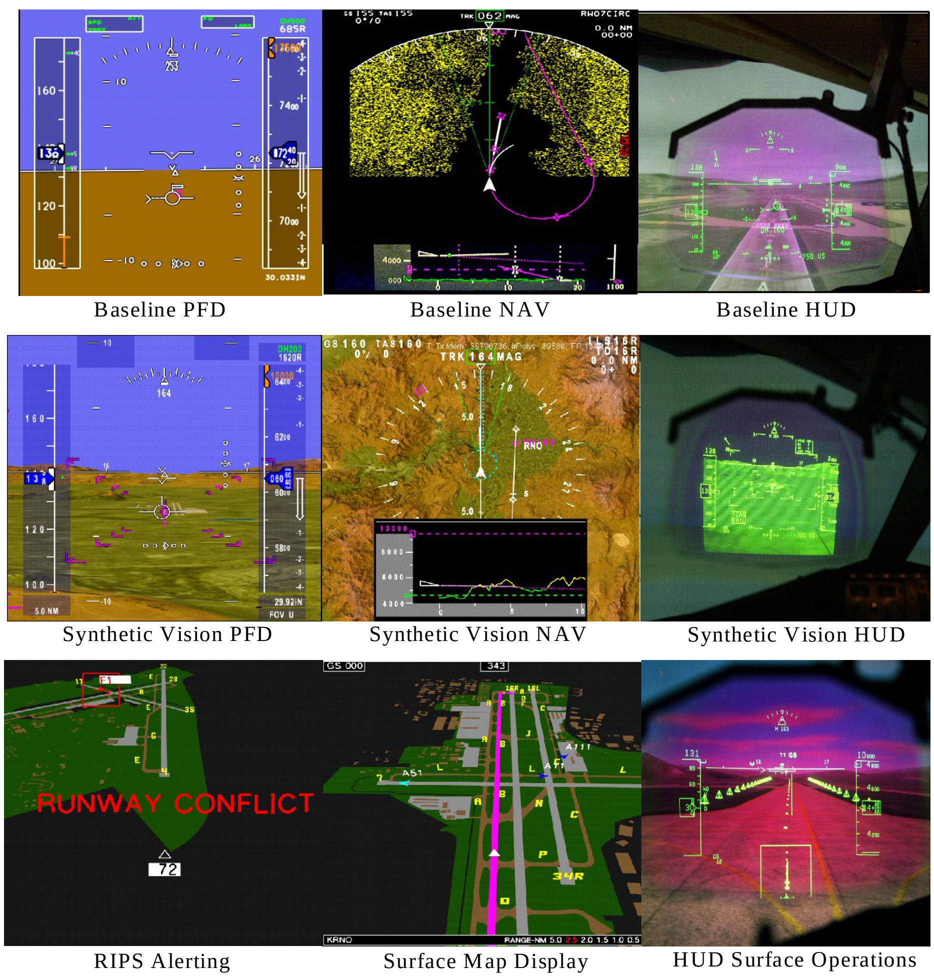


Figure 2. *Experimental Display Concepts*

IN-FLIGHT PLANNING AND INTELLIGENT PILOT AIDS FOR EMERGENCIES AND NON-NOMINAL FLIGHT CONDITIONS

A.R. Pritchett and **V.V. Kalambi**
Georgia Tech, Schools of Aerospace Engineering and Industrial and Systems Engineering

A commercial flight plan comprises a series of turns and climbs or descents defined by headings or waypoints, and speed and altitude constraints at each. Situations do occur in-flight where the pilot must re-plan, inflight, the lateral and vertical profile of the remainder of the flight. For example, a 'non-nominal' condition such as a thunderstorm may require re-routing; less frequently, an emergency situation may require an immediate landing at the nearest airport. . The objective of this research was to see how pilots perform in-flight planning by observing the planning behavior of pilots in non nominal and emergency conditions arising in the last 15-30 minutes of flight. The impact of autoflight systems on planning was also examined, including notional systems with the capability of automatically generating a flight plan.

Results from a medium-fidelity flight simulator experiment with airline pilots showed that the autoflight systems did not have a significant impact on the replanning task. Instead, the specific scenario showed more of an effect on the primary performance measures of time of flight and distance flown. Interesting trends of lateral and vertical navigation were also seen, together with sometimes unconventional use of the autoflight systems. Pilots always tended to go for the most direct route possible when given discretion. Pilots did not verbally express any distinction between emergency and non-nominal flight conditions, however, the effect of these flight conditions was seen when the planning performance measures of time of flight and distance flown were analyzed. Most pilots were quite aggressive with their plans in terms of speeds and descents at higher altitudes but maintained shallow turns onto final approach.

Pilots favored the use of the automatically generated plan. From the experiment results it was determined that automatic flight path generation would be beneficial to the task of in flight re-planning and would only serve to reduce the workload in high workload emergencies. However, it is imperative that, for such a system to be useful, it should have the ability of considering a number of contextual factors simultaneously, including real time access to information about the immediate context, including traffic, weather and terrain.

SIMULATING GLASS COCKPIT DISPLAYS IN A GENERAL AVIATION FLIGHT ENVIRONMENT

Robert W. Proctor
John P. Young
Richard O. Fanjoy
Robert G. Feyen
Nathan W. Hartman
Vishal V. Hiremath
Purdue University
West Lafayette, Indiana

Development of a research platform that replicates the basic flight functions of a light, general aviation aircraft is described. This involved retrofitting an actual aircraft cockpit with computer displays to emulate an aircraft environment. The hardware and software used in this research platform are described, as well as issues and problems regarding implementation and use in research.

Introduction

Progress in computer and electronic technology has dramatically impacted the capabilities and design of modern avionics. The cockpits of most commercial and military aircraft of today bear little resemblance to those of 30 years ago. Recent advances in avionics have embraced the concept of combining the equipment, functions, and displays from various flight information and navigation systems into one or two highly integrated units. Large commercial and military aircraft have evolved in their use of electronic and computer technologies to the point where most modern aircraft utilize Electronic Flight Instrument Systems (EFIS), commonly referred to as "glass cockpit" technology. The electronic display units of glass cockpits are flexible and can present a vast array of information in different ways over which the pilot has control (Billings, 1997). Because these electronic displays are more user-friendly than traditional cockpit displays, the switch to EFIS has improved system reliability and safety.

This transition to EFIS has not been without its drawbacks, however. Analysis of recent airline accidents has demonstrated that there are many human factors involved in transitioning to and using such display technology (e.g., Kaber, Riley, & Tan, 2002). The pilot's mental workload may be reduced during routine flight, but when an unusual condition occurs or a flight plan needs to be altered, the workload may increase dramatically (Ishibashi, 1999). Due in part to a reduced role of the pilot in the aircraft control loops, during the times of high workload the pilot's decision making and performance may suffer from inadequate situation awareness, or comprehension of the factors involved in the current situation.

Although larger air carrier and corporate jet manufacturers' aircraft have incorporated EFIS into their flight decks, little has been done in the General Aviation (GA) market until recently (e.g., Williams, Yost, Holland, & Tyler, 2002). Previous GA aircraft did not employ glass cockpits due to expense, immature miniaturization technology, and a lack of available space in the cockpit. However, a transition to glass cockpit technology in GA is beginning to occur due to the need to present more and more information in formats that a pilot can effectively utilize. To make small aircraft more accessible for greater numbers of pilots and to ease the impact of small privately owned aircraft on the air traffic system, the FAA and NASA have initiated the Small Aircraft Transportation System (SATS) program for the development of highly integrated and advanced technologies for GA.

Given the human factors problems associated with the switch to the glass cockpit in commercial and military aviation, where the pilots are professionals with extensive experience and training, it should be apparent that the potential threat to safety with introduction of the glass cockpit in GA is great. Reduction of this threat requires systematic human factors analyses to guide the design and development of the interfaces. Although glass cockpits are currently being introduced by GA aircraft companies, little thought seems to have been given to human factors concerns. Features of commercial glass cockpits are being incorporated directly without consideration of their suitability for relatively inexperienced GA pilots; little standardization is evident across different company's displays; and training in use of glass cockpits is minimal and unsystematic. These are a few among several human-factors issues that must be addressed for increasingly advanced EFIS technology to be implemented successfully in smaller GA aircraft (Feyereisen & Cundiff, 2001).

Of immediate concern is the need to identify and explore problems associated with the transition of general aviation pilots from “gauge” (analog) style flight information display technology (on which they were initially trained) to highly integrated glass cockpit formats. For example, Jones (2004), in describing his first look at the Garmin 1000 equipped Cessna 172, states, “I have to admit, it is an impressive panel. Usually reserved for the high-end airplanes and airliners, this panel is a video game junkies dream come true. Although it made me a little uncomfortable not having the old steam gauge instruments sitting front and center” (p. 2). In the future, when pilots have been trained from the beginning with glass cockpit displays that have been designed to thoroughly address human factors concerns, there should be a major improvement to aviation safety. However, in the near term, problems associated with transitioning pilots who were initially trained using gauge displays could easily override these benefits if not dealt with satisfactorily.

Because of the need for systematic investigation of human factors issues associated with the adoption of glass cockpit technology in GA, we have begun to conduct an interdisciplinary research program to examine these issues. As part of this process, we developed a research flight simulator that replicates the basic functions of a light, general aviation aircraft, and allows us to control many aspects of the cockpit displays while measuring several aspects of pilot performance. In this paper, we describe the hardware and software used in this research simulator, issues and problems encountered in simulator development, and the decisions made at each point. We also outline the capabilities and limitations of the simulator, and discuss plans for future research using this device.

Simulator Development

Our goal was to construct a research platform that would provide maximum realism, or ecological validity, within a budget of approximately $6,000 (see Table 1). An optimal system would allow us to vary and control design and flight parameters, and to record various the measurements of pilot performance.

To start with, we salvaged a KingAir cockpit shell for use with the simulator equipment (see Figure 1). The front 15 feet of the fuselage had been previously separated from the rest of the aircraft; we cleaned and stripped out this section. We then added new flooring and wall paneling, and supports for the simulator equipment. The nose compartment was converted into the housing for the simulation computers and other related equipment. The equipment and costs are summarized in Table 1. We selected Dell Model GX270 Pentium 4 computers operating at 3.0 GHz, with 1 GB RAM and a 128 MB

Table 1. *Equipment List and Costs for Simulator*

Equipment Item	Vendor	Quantity	Cost
OptiPlex GX270 Small MiniTower—Intel Pentium 4 Processor 3.00GHz, 1GB RAM	Dell	2	$2,310
128MB nVidia GeForce FX 5200 Graphics Card	Dell	2	$206
1224L 12-inch LCD Desktop Touch Monitor	Dell	1	$540
MicroTouch 17-inch CRT Touch Monitor	Dell	2	$869
G90fB 19-inch PerfectFlat Black CRT Monitor	Dell	2	$416
X-Plane	GraphicSim	1	$50
Throttle Panel, Push-button/Toggle switch module, Landing gear module, Autopilot module, etc.	Goflightinc.com	1 each	$1080
Yoke, Flight Controls, Pedals, Sidesticks, etc.	CH Products	1 each	$390
*GL Studio	Distributed Simulation Technology	1	$4,500

* GL Studio was not included in the original price estimate listed in the body of the paper

nVidia GeForce graphics card, to accommodate the large amount of memory needed to display the graphics. As currently set up, one computer handles the external view and the other computer handles the in-cockpit view, and each computer is connected to a pair of monitors via a split VGA adapter cable.

Figure 1. *General aviation flight simulator.*

For the external view, two 19-inch flat CRT monitors are mounted outside the shell just outside the forward windows. The external image is spread across the two monitors to present the simulated outside world visible from the cockpit (see Figure 2). We investigated the use of a projection system to provide a more realistic panoramic view, but the cost required to overcome hurdles in the projection screen resolution and brightness was too high. Initial reports by users have been favorable, indicating that despite some initial "tunnel vision" (due to the lack of view through the side windows) and the small field of view, the CRT displays are sufficiently realistic for the users to report that, after several minutes, they "forget" they are looking at computer monitors and instead treat them as the "real world" view out of the **front windows.**

Figure 2. *Displays of simulated external view and instrument panel.*

For the in-cockpit view, two 17-inch "touchscreen" monitors were fitted into the instrument panel, one on each side of the cockpit, to present the displays seen by the pilot (see Figure 2). The right display is included for use by either a co-pilot or by an experimenter. We selected touch screens to allow us the option of exploring touch-based manipulation of the displays and controls in the general aviation setting. A third 12-inch touchscreen monitor (not shown) was also purchased to explore alternate methods of displaying and entering flight management data.

We opted to go with commercially available flight control software rather than develop our own software. After evaluating two mainstream flight simulator software packages used in the gaming community (Microsoft Flight Simulator and X-plane), we decided on X-plane because of its superior graphics quality and data collection capability. One significant advantage of X-plane is a feature called PlaneMaker, which allows the user to design custom cockpits and instrument layouts. The instruments for the panel can be selected from a database available in PlaneMaker and placed in the locations of the users' choosing. Virtually every flight parameter in X-plane can be recorded, which is a very useful feature for research. X-plane provides a comprehensive list of flight parameters, any of which are selectable by the user. Although X-plane's data recording capabilities appear sufficient for the time being, we also have installed LabView data acquisition software to support more advanced data collection and analysis.

We chose flight controls by CH Products primarily on the recommendations of the manufacturer of the X-plane software. Installation was simplified by their plug and play capability with X-plane. We planned originally to implement controls on both sides of the cockpit, as in an actual aircraft. However, X-plane does not allow two controls to operate simultaneously, which restricts control to one pilot. Thus, we decided to purchase two sidesticks and one yoke so that we could simulate as many aircraft types as possible. Although the realism of the sidesticks is enhanced if they provide force feedback capability, we did not choose to include this capability due to our current budget limitations.

Another major limitation of the X-plane software is that two different cockpit displays cannot be displayed on the two different monitors. For example, it is not possible to run the primary flight display on one monitor and a multi-function display on the other. This is because X-plane does not allow the cockpit to be stretched across both screens so that, instead, both screens show the same image. We hope to overcome this hurdle with GL Studio, the software that we purchased to design custom instrumentation. GL Studio is a reasonably priced software tool that can be used to design glass cockpit instrumentation that can then be linked to the flight simulation software. Given our interest in exploring alternative instrumentation designs, the capability to design our own instrumentation and layouts was a crucial addition to our simulator platform.

Research Agenda

The advanced flight display platform described in the preceding section was developed to address a wide variety of flight instrumentation and training issues. Many of these issues have been identified in the body of literature that has evolved in response to nearly two decades of glass cockpit operations in commercial service. These issues have attracted new interest with the recent development and implementation of glass cockpit instrumentation in general aviation aircraft. The focus of our research effort is to identify relationships between mental models of pilots and advanced instrumentation designs. A primary objective is to identify displays that effectively support pilots' mental models and permit intuitive responses to environmental inputs.

The first stage of this study will be a two-pronged effort to map mental models used by pilots of varying experience levels and to identify differences between current glass cockpit designs used in general aviation aircraft. These steps will be followed by an investigation of cognitive "disconnects" that occur during glass cockpit flight operations. Finally, investigators will attempt to modify display aspects and training curricula to foster improvements to pilot performance in a glass cockpit environment.

The research platform we have detailed provides an inexpensive, yet robust resource to identify key aspects of pilot mental models and flight display efficiencies. It is not, however, a high-fidelity flight simulator that approaches FAA certification standards. Data collected with the platform will more accurately reflect discrete display and performance aspects, and should not be generalized to a complete and accurate flight environment without additional study. Initial subjects will constitute samples of convenience from a general university flight student population, but follow on efforts may address a wider, general aviation pilot population.

We are currently conducting a preliminary experiment that will allow us to more completely define the simulator's capabilities and limitations. This experiment basically explores the difference in recovery times of pilots when they are flying an analog cockpit display versus a glass cockpit display. We are examining the effects of changing instrument design on performance, for example, changing a dial indicator to a vertical tape indicator. Students who recently completed general aviation pilot training will fly a scenario that requires recovery from an unusual attitude, that is, from a situation in which the aircraft is in an abnormal position with relation to the horizon (e.g., being very high nose up). Recovery time will be measured from the instant that the pilot first receives indication of the unusual attitude to when s/he returns the aircraft back to wings level flight and cruise airspeed configuration. The results of this experiment will allow us to become familiar with all details of the simulator and provide an initial step toward accomplishing our longer-range research objectives.

References

Billings, C. E. (1997). *Aviation automation: The search for a human-centered* approach. Mahwah, NJ: Erlbaum.

Feyereisen, T. & Cundiff, C. (2001). Visual cueing and control for general aviation application. *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, v 1, pp. 5B11-5B15.

Ishibashi, A. (1999). Situation awareness in the automated glass-cockpit. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 3*, p. III-710 - III-714.

Jones, N. (2004). Chesapeake airport open house. *The Flying Wire* (Monthly publication of EEA Chapter 339), July, 04, p. 2. (Downloaded January 31, 2005, from http://homepage.mac.com/chapter339/.cv/chapter339/Public/2004-07.pdf-link.pdf)

Kaber, D. B., Riley, J. M., & Tan, K.-W. (2002). Improved usability of aviation automation through direct manipulation and graphical user interface design. *International Journal of Aviation Psychology, 12*, 153-178.

Williams, K. W., Yost, A., Holland, J., & Tyler, R. R. (2002). Assessment of advanced cockpit displays for general aviation aircraft: The Capstone Program. *FAA Office of Aviation Medicine Reports. DOT/FAA/AM-02/21, Dec 2002, 1-35. US: Aviation Medicine.*

EFFECT OF AIR TRAFFIC CONTROLLER TASKLOAD AND TEMPORAL AWARENESS ON TASK PRIORITIZATION

Esa M. Rantanen and **Brian R. Levinthal**
University of Illinois at Urbana-Champaign
Savoy, IL 61874

This paper describes an experiment that was conducted to provide an empirical foundation for estimation of parameters for air traffic controller performance modeling efforts presently pursued within the NASA DAG-TM CE-6 model development. The focus of the work was the task prioritization scheme used in these models. A total of 11 retired FAA controllers and supervisors assigned to the FAA Technical Center volunteered to participate in the experiment. A part-task experimental simulation that presented the participating controllers with several simultaneous tasks in four quadrants, or panes, on a single display was used. Only one pane and typically one task could be viewed at a time. This allowed for measurement of controllers' attention to each task. All events unfolding in the experimental scenarios and controllers' actions were recorded and timed as well. From these data, several dependent variables were derived, focusing on the temporal aspects of controllers' performance and their prioritization of simultaneously available tasks. The results indicate that taskload was manipulated successfully and resulted in measurable differences between experimental conditions in both taskload and performance, the latter evinced by the time elapsed in a window of opportunity for a given task before action was taken on it as well as time remaining in the window of opportunity when action was completed. However, it appears that either the controllers were not aware of these temporal features of their tasks or that other factors dominated their prioritization decisions. Task prioritization may hence be driven by task characteristics that are categorical rather than continuous and quantifiable.

Introduction

The intricacies of control of complex and dynamic systems are particularly well illustrated in the nation's air traffic control (ATC) system. The importance of an up-to-date mental model of the traffic situation to the controller is self-evident, as are the temporal demands of the controllers' task. Anticipatory behavior of air traffic controllers, however, is not overt: Anticipation is not an end in itself, it is seldom expressed in verbal communications, and may not result in directly observable behavior (Boudes & Cellier, 2000). Yet, accurate anticipation lies in the core of successful control of air traffic and the performance of a controller by allowing early detection of conflicts (i.e., two aircraft coming closer to each other than a minimum separation required) and formulation of conflict-free traffic flows. A controller who fails to anticipate the development of traffic situation has already 'lost the picture' and is forced from proactive into a reactive mode of behavior, rapidly increasing his or her stress, workload, and propensity for unrecoverable errors.

The temporal dimensions in controllers' tasks are also likely to gain in significance with the introduction of automation applications in ATC. One large-scale effort to increase the National Airspace System (NAS) capacity is the NASA Distributed Air/Ground (DAG) Traffic Management (TM) concept of distributed decision-making. The goal of DAG CE-6 (Concept Element) is to integrate the controller DST with data link to minimize lags/delays while providing controllers with as much flexibility in options as they have today. However, substantial qualitative differences in the working methods and practices of the controllers are to be expected (c.f., Hopkin, 1995; Wickens et al., 1997; Wickens et al., 1998). These differences may in turn have important impact on the controllers' performance and workload, potentially quantifiable by the temporal characteristics of their tasks and the way they are carried out.

In addition to the importance of temporal performance of air traffic controllers, time may offer a useful domain for research of a multitude of human factors aspects. All scientific research of mental models and subsequent engineering applications are dependent on methods of measurement (Chapanis, 1959). Apart from the relevance of time to anticipatory behavior in control of dynamic systems, it offers attractive methods for the measurement of covert mental models. Time has a long history as a means to investigate cognitive processes, manifested by extensive reaction time research. Timing data (e.g., response times) are relatively easy to obtain under both experimental and naturalistic conditions, and time is a variable that is common to the human, the task, and the environment. Time offers thus a common unit of measurement of human performance in the context of the task, and can be used to infer the goodness of the temporal dimension of the operator's mental model of the task or system being controlled. Grosjean and Terrier (1999) defined temporal awareness as a "representation of the situation including the recent past and the near future," (p. 1443) echoing definitions of mental models (e.g., Rouse & Morris, 1986) and situation awareness (e.g., Endsley, 1995). In an experiment mimicking control of three simultaneous processes (simulated production lines) Grosjean and Terrier (1999) discovered that subjects who had developed good temporal awareness made fewer errors, prioritized their work more effectively, and managed their rest periods better than those with poorer temporal awareness. Temporal awareness was thus found to be a good predictor of performance.

Task network models use human/system task sequence as the primary organizing structure and hence appear as particularly suitable approach to modeling air traffic controllers' jobs, which consist of many tasks with varying degrees of dependency. As all tasks and subtasks unfold in time, it may be hypothesized that their successful management is primarily a temporal task and the controller's performance is predominantly determined by his or her time management skills and the goodness of his or her temporal awareness of the situation. Time is hence an attractive variable for investigating the interactions of ATC task load and controller performance as well as a congenital parameter in task network models. The purpose of this research was to provide an empirical foundation for estimation of parameters for air traffic controller performance modeling.

Method

Participants

Participants for this study were recruited among retired FAA controllers and supervisors assigned to the FAA Technical Center's Human Factors Research and Development Laboratory (HFRDL) at Atlantic City International airport, NJ. A total of 11 volunteers participated in the experiment. All participants were male, with a mean age of 55.64 years (SD = 9.1), ranging from 38 to 66 years. All were also very experiences in a variety of ATC facility types with a mean experience as a controller of 23.45 years (SD = 6.67), ranging from 11 to 33 years.

Apparatus

The experimental apparatus was a custom-built ATC simulator. The simulation program was written in C++ and ran on two laptop computers with 14-inch TFT displays and 1024 x 768 –pixel resolution. The simulator mimicked the display system replacement (DSR), including data link (DL) capability, allowing for accurate timing of participant interactions with the DL interface. A regular mouse was provided for moving between tasks (as described below) and control inputs.

Experimental Task

The experimental task mimicked the job of air traffic controllers. The participants viewed air traffic scenarios on four separate quadrants, or panes, on a single computer display. The scenarios could be viewed only one at a time by moving a cursor to the desired pane. This task balanced the requirements of realism and experimental control and it allowed for accurate measurement of times of the different events unfolding in the experimental scenarios as well as timing of the participants' actions in response to them. Six subtasks modeled in the NASA CE-6 modeling effort were selected for the experiment: (1) receive handoff, (2) initiate handoff, (3) transfer communications, (4) respond to DL request to change altitude, (5) perform conflict resolution, and (6) perform metering.

Design

Independent variables. The primary independent variable was taskload, which was manipulated through several other variables over which the experimenter has complete control. It should be noted, however, that control over these variables was constrained by the participants' actions after the onset of the experiment, that is, the eventual sequence and timing of the tasks depended on individual participants' different time management skills and strategies as well as other individual performance differences. Time required (TR) was manipulated primarily by differential difficulty of conflict situations, based on findings of Rantanen and Nunes (in press). Pilot testing revealed a mean time required for participants to use the datalink system's flyout menus to communicate altitude, speed, or heading clearances to pilots, respond to downlink requests, and initiate and receive handoffs. Time available (TA) consisted of the individual windows of opportunity (WO) for each task encountered per trial. In each trial, certain windows of opportunity overlapped reliably, regardless of individual difference in performance, as a result of the discrete trial onset times. For example, the WO for receiving flights from handoff would not vary across participants because it was related directly to trial onset time and the initial speed of the flights, while the extent of the WO for conflict resolution, DL responses, or resolving metering violations would be subject to individual differences. The ratio of time required and time available was the basis of the definition and computation of nominal taskload. By nominal we mean that it was calculated a priori, at the outset of the scenarios. A total of 31 scenarios were created and used to form a total of 20 experimental scenarios: 8 high taskload conditions, 8 low taskload conditions, and 4 transitions.

Design. The basic design of the experiment was a 3 (taskload, Low, Transition, High) x 2 (order Lo-Tr-Hi, Hi-Tr-Lo) x 2 (replicates) factorial design. In the subsequent analyses, however, only low and high taskload conditions were considered, the transitional scenarios split between the two conditions. Four scenario files, one file per quadrant (pane) on the display, started the experimental blocks. At the end of each scenario, a new scenario files filled the pane. An experimental block was comprised of 4 (panes) x 5 scenarios, which followed each other in a seamless sequence. Three levels of taskload were included in each block: two scenarios per pane of high taskload, two scenarios per pane of low taskload, and one transition scenario per pane. The order the scenarios were presented was balanced so that each participant encountered a block that started with low taskload (first two scenarios per pane) and ended with high taskload (the last two scenarios per pane) as well as a block in which the scenarios with different taskload were presented in an opposite order (starting high and ending low).

Dependent variables. To derive the objective performance metrics, a number of actions were timed and recorded for each task. These timed actions were used to derive a number of dependent variables for the purposes of this research. The elapsed time from opening of WO at the time the task was performed (time to first action, or TFA) was calculated by subtracting the time the WO opened for the task from the first action on the task (e.g., mouse click on a flyout menu). Note that this value may also be negative if the task was performed before the WO opened. The time remaining in the WO after completion of the task (TRm) was calculated by subtracting the time of the last action on the task from the time the WO for that task closed. It was hypothesized that good performance would be manifested by prompt actions in tasks (small TFA values) and ample time remaining after completing the task (large TRm values).

Results

The experimental simulation program recorded all events and actions taken by the participants into a text file in a form of a time line. The data were then processed by another program for reduction. This program read the timeline and reorganized the data into an output file so that tasks were entered in rows every time the participant did something about them (including looking at a task, or 'dwelling' in it), plus other measures pertinent to that particular task.

We wanted to determine the actual taskload as influenced by the participants' control actions and strategies, as we anticipated the actual taskload to be different from the nominal one determined from the outset of the experiment. An index of taskload (actual taskload, TL_A) was provided by the following formula:

$$TL_A = \frac{n(TR_{avg})}{TE} \quad (1)$$

where n is the total number of tasks present in an epoch and TR_{avg} the average time required to perform these tasks. The TE is the duration of the epoch, in this case 300 s (5 min).

It is acknowledged that many tasks had zero time required to perform them, for example, acceptance and initiation of handoffs and transfer of communication only required a single mouse-click. Furthermore, it is clear that physically performing the task, by keyboard entries or clicking through menus with a mouse, only constitutes a small fraction of the total time required to perform the task, that is, the overt actions do not reveal planning and decision-making processes, which almost certainly require most of the controller's time. Nevertheless, multiplication of the time required by the number of tasks compensate to some degree the very short (i.e., 0) performance times in an epoch, and indeed this index showed clear differences between the different taskload conditions (Figure 1). The differences between taskload were significant (two-sample t-test, $p < .05$) for all but the transition epoch.

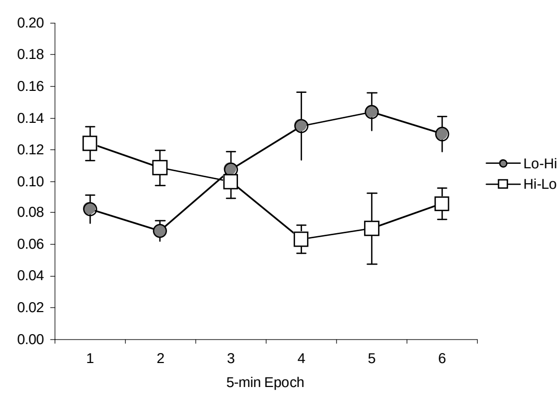


Figure 3.1. Mean actual taskload index values (by eq. 1) by taskload condition and 5-minute epochs; the switch from low to high and high to low taskload condition about 10 minutes into the block is apparent.

Task prioritization was analyzed by observing which tasks were performed before others when all were 'available' simultaneously, that is, the WO for performing the tasks were open at a same time. Specifically, the probabilities a given task was chosen to be performed first among a number of simultaneously available tasks were derived by the following method:

(1) Divide the experimental block in time 1-minute epochs. This epoch duration was somewhat arbitrary, but its minimum was determined by the necessity to have more than one task 'available' within it and its maximum by the notion of simultaneity. The average number of tasks in the 1-minute epochs was 6.45, with a range from 2 to 11.

(2) For each epoch and tasks in the epoch, the actions taken by the controller were recorded (essentially, the first action a controller took on a task).

(3) Based on the first-action times, tasks within an epoch were ranked (1st, 2nd, etc.)

(4) These data were then sorted by task and each task pair was analyzed separately, counting the times a task in a pair was acted on before the other task(s) in the pair within an epoch.

(5) These counts were then summed across participants and experimental blocks, and the proportion of times one task in a pair was acted on before the other was calculated.

(6) The above procedure was repeated for another set of 1-minute epochs, offset by 30 s from the above, to maximize the number of task pairings. This could be seen as a resampling technique, and the combined results improve the reliability and accuracy of the probability estimates.

The results of this analysis are depicted in Table 1. Taskload condition appears to have had only minimal impact on prioritization between tasks in pairs. However, before interpreting these results particular limitation of this analysis must be observed: this method considered the tasks separately, that is, whether a particular task was performed first or not within the 1-minute epoch, and assigned the task into a group based on the given variable value (TFA or TRm) for that task only. In reality, however, the tasks were not independent but considered by participants relative to each other. To determine whether TFA of TRm of each task in a pair was a factor in the participant's choice of task to be performed first, other methods of analysis must be employed.

Table 1. *Probabilities (proportions) a given task was performed before another task when both were available (i.e., their WOs were 'open') simultaneously. Key to the task acronyms: CR = Conflict Resolution, DL = Downlink request (climb/descent), FR = Frequency Change, IH = Initiate Handoff, MV = Metering Violation, RH = Receive Handoff.*

Taskload	Proportion	
Task Pair	CR/DL	DL/CR
Low	0.592	0.408
High	0.471	0.529
	CR/MV	MV/CR
Low	1.000	0.000
High	0.714	0.286
	CR/RH	RH/CR
Low	0.306	0.694
High	0.331	0.669
	DL/FR	FR/DL
High	1.000	0.000
	DL/RH	RH/DL
Low	0.275	0.725
High	0.334	0.666
	FR/IH	IH/FR
Low	0.400	0.600
High	0.125	0.875
	RH/FR	FR/RH
Low	1.000	0.000
High	0.833	0.167
	RH/MV	MV/RH
Low	0.875	0.125
High	0.839	0.161
	DL/MV	MV/DL
Low	0.750	0.250
High	0.500	0.500
	MV/RH	RH/MV
Low	0.188	0.813
High	0.250	0.750

As was discussed above, it was of interest to examine whether the participants' temporal awareness, that is, awareness of the TFA or TRm of each task at hand (i.e., tasks with simultaneously open WOs) played a role in their decisions to prioritize one task over another. To do this, we considered tasks in pairs, as was done in previous task prioritization analyses. In this analysis, however, we contrasted the TFA and TRm values of each task in a pair according to the eventual priority given to a task. Specifically, we hypothesized that it the controller was aware of the time elapsed in the WO (TFA), he or she might perform the task with longer TFA first and the task with a more recently opened WO second. Hence, the hypothesis may be operationalized as

$$TFA(1) > TFA(2) \Leftrightarrow TFA(1) - TFA(2) > 0 \qquad \text{(H1)}$$

Another hypothesis was that if the participants were aware of the impending *closing* of a WO, that might have contributed to a sense of urgency and a task with a shorter TRm value would be performed first. Specifically,

$$TRm(1) < TRm(2) \Leftrightarrow TRm(1) - TRm(2) < 0 \qquad \text{(H2)}$$

We tested these hypotheses for task pairs of conflict resolution (CR) and datalink request (DL) as well as CR and receiving handoff (RH) by calculating the proportions of positive and negative outcomes of the above hypotheses (see Table 2 below).

Table 3.15. *Proportions of the positive and negative outcomes from the hypotheses H1 an H2 as stated above. However, the results are not only mixed (i.e., split between positive and negative) but actually opposite to the hypotheses. In both task pairs, TFA had a higher proportion of negative values than positive, and the TRm a higher proportion of positive values than negative*

Task Pair	Variable	N Pos	N Neg	% Pos	% Neg.
CR/DL	TFA(1)–TFA(2)	81	166	32.8	67.2
CR/DL	TRm(1)–TRm(2)	125	77	61.9	38.1
CR/RH	TFA(1)–TFA(2)	98	352	21.8	78.2
CR/RH	TRm(1)–TRm(2)	186	81	69.7	30.3

The results did not confirm the hypotheses. As a matter of fact, they were opposite to what was hypothesized in that the participants seemed to perform tasks with more recently opened WOs before tasks that have been available longer, and tasks with more time before closing of their WOs before more urgent tasks by the same measure.

We also performed an ANOVA to see whether the above hypotheses differed between task priority and taskload, with taskload level and task priority as factors, plus their interaction in the model. For CR/DL and TFA, neither of the main effects was significant, that is, the difference in TFA values between the tasks in the pair did not differ significantly between taskload conditions or between task priorities. The interaction between these factors was significant, however, $F(1, 243) = 14.87$, $p < .001$. For TRm, task priority was significant, $F(1, 198) = 638.55$, $p < .05$, but no other factors or interactions. Analysis of the CR/RH task pair yielded similar results; for TFA, only the interaction between task priority and taskload was significant, $F(1, 446) = 4.84$, $p < .05$, and for TRm, there were no significant results.

Discussion

The combined results from this study suggest that task prioritization may be driven by task characteristics that are categorical rather than continuous and quantifiable. Support for this conclusion is provided by the very different trends in TFA for the three different tasks analyzed, conflict resolution, receiving handoffs, and responding to downlinked requests. Of these, conflict resolution was clearly the most difficult task, as well as the most important. The difficulty of detecting conflicts as well as the time required to construct and implement resolutions to them probably made this task more vulnerable to influences of workload and time pressure than simpler tasks. It must also be remembered that accepting handoffs is, in addition to being a quick and easy task to perform, a prerequisite to subsequent control of the flight (e.g., to implement conflict resolution) and hence the average prioritization between conflict resolution and receiving handoffs is inherently biased towards the latter.

Another aspect worth considering is the nature of the analyses and differences between experimental simulations and realistic situation in operational ATC. Statistics (i.e., minimization of probabilities of both Type I and II errors) is dependent on sufficiently large number of observations, which necessitates aggregation of observations across individual participants and experimental blocks. Yet, even in relatively constrained task environments such as our experiment these observations exhibit substantial variability. For example, aggregation of conflict resolution tasks and receiving handoffs as was done here did not consider the often unique characteristics of each of these instances. Parsing the data according to such characteristics, however, would severely limit the number of observations available for analysis and undermine the reliability of the results. This is a classical 'Catch-22' situation for which the only remedy is to collect much more data over extended periods of time.

Finally, large differences in performance of individual participants should not be overlooked. These differences were statistically highly significant in almost all analyses we performed and bespeak of inherent variability in working techniques, strategies, and performance of individual controllers working on the same tasks.

Acknowledgments

This work was supported under National Aeronautics and Space Administration and Micro Analysis and Design Contract NASA-MAD-7400.005.01. Parimal Kopardekar from NASA and Ken Leiden from MAAD served as technical monitors. Views expressed herein are those of the authors and do not necessarily represent official NASA, FAA, or MAAD positions. We express our appreciation to Ben Willems and Pamela Della-Rocco at the FAA William J. Hughes Technical Center for their invaluable assistance in securing participants and venue for running the experiment and the individual participants for volunteering their time and expertise for this study. Special thanks are due to Sharon Yeakel for programming the experimental simulator and data processing tools.

References

Boudes, N., & Cellier, J.-M. (2000). Accuracy of estimations made by air traffic controllers. *International Journal of Aviation Psychology, 10*(2), 207-225.

Chapanis, A. (1959). *Research techniques in human factors*. Baltimore, MD: Johns Hopkins University Press.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Grosjean, V., & Terrier, P. (1999). Temporal awareness: Pivotal in performance. *Ergonomics, 42*(11), 1443-1456.

Hopkin, V. D. (1995). *Human factors in air traffic control*. Bristol, PA: Taylor & Francis.

Rantanen, E. M., & Nunes, A. (In press). Hierarchical conflict detection in air traffic control. *International Journal of Aviation Psychology*.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100*(3),349-363.

Wickens, C. D., Mavor, M., & McGee, J. (1997). Flight to the future: *Human factors of air traffic control*. Washington, DC: National Academy of Science.

Wickens, C. D., Mavor, M., Parasuraman, R., & McGee, J. (1998). *The future of air traffic control*. Washington, DC: National Academy of Science.

THE IMPACT OF SECTOR CHARACTERISTICS AND AIRCRAFT COUNT ON AIR TRAFFIC CONTROL COMMUNICATIONS AND WORKLOAD

Esa M. Rantanen, **Philip W. Maynard** and **Deniz Özhan**
University of Illinois at Urbana-Champaign, Aviation Human Factors Division
Savoy, Illinois, USA

Means of communication between pilots and controllers is one of the fundamental principles of air traffic control (ATC). Consequently, air-ground communications will both reflect the taskload imposed on the controller as well as drive the workload experienced by the controller. Therefore, analysis of ATC communications could potentially reveal a very rich and detailed picture of the demands placed on a controller in a given sector and traffic situation. This paper reports analysis of ATC voice data obtained from three different sectors at the Indianapolis air route traffic control center (ZID ARTCC). The main purpose of this analysis was to examine how different sector characteristics and the busy and slow periods within the sectors differed explicitly in terms of pilot-controller communications and implicitly in terms of controller taskload and workload. Measures derived from the voice data were also compared to metrics reflecting ATC sector complexity that were derived from the output of the objective activity and taskload assessment program, POWER, developed by the FAA.

Introduction

Mental workload in frequently cited as the most critical characteristic of air traffic controller's task (e.g., Hopkin, 1995; Wickens, Mavor, & McGee, 1997; Wickens, Mavor, Parasuraman, & McGee, 1998). Mental workload experienced by controllers is frequently cited as a limiting factor to the capacity of the entire national airspace system (NAS) and evaluation of new technologies and forms of automation is often focused on their impact on controller workload in particular. Mental workload, however, is a complex and multidimensional theoretical construct, influenced by numerous interacting factors (Meshkati, 1988; Vidulich, 2003), and with a substantial subjective component. Hence, workload is covert, individually experienced by controllers and not directly measurable.

Controller workload could, however, possibly be inferred from other, overt, aspects of controllers' work. For example, there has been much research activity to quantify taskload by measurable characteristics of traffic in an ATC sector or by the infrastructure of the sector itself, collectively knows as sector or traffic complexity or dynamic density. Much of air traffic controllers' work also involves spoken communication. Presently, virtually all control actions must be communicated to pilots via voice radio. Hence, voice communications are intuitively and unsurprisingly an attractive method for examining controller workload. However, little research has been done to validate and quantify the putative relationships between sector complexity and workload on one hand, and controller communications and workload on the other. The purpose of the research reported in this paper was to examine the relationship between various sector complexity measures and ATC communications, and thus attempt to bracket controller workload.

ATC Complexity and Workload

Dynamic density is used in a variety of contexts in the literature and does not necessarily correspond to a single metric, but Laudeman, Shelden, Branstrom, and Brasil (1998) and Sridhar, Sheth, and Grabbe (1998) have reported an equation for this construct. The index sums nine specific variables, each multiplied by a weight derived from regression analysis of controller activity data and subjective ratings. No definition was provided for traffic density, however. Another complexity metric, risk index, is an index of collision risk (Knecht, Smith, & Hancock, 1996) and it has also been referred to as dynamic density (Smith, Scallen, Knecht, & Hancock, 1998). It is derived from two directly measurable variables, (1) number of aircraft at a given altitude, N, and (2) distance from the *i*th to the *j*th aircraft, *d*ij Other complexity metrics include predictive workload measures based on work done at NASA Ames research center by Chatterji and Sridhar (2001), and work by Wyndemere, Inc. (1996; see also Pawlak, Brinton, Crouch, & Lancaster, 1996 and Pawlak & Brinton, 1996). Summaries of the many metrics are provided by the FAA (2000) and Kopardekar and Magyarits (2002).

No theoretical foundations for measurement of these constructs could be established from the ATC research literature, as such were not provided. Instead, it often seems to be the case that validity of inferences made about covert, not directly measurable constructs is based only on the authors' proclamation that by measuring A (a directly measurable variable) they were in fact also measuring B (a covert, only indirectly measurable variable). Hence, much research remains to be done to create and validate a theoretical framework for establishing rigorous and reliable connections between directly measurable variables and indirect constructs of interest.

ATC Communications and Workload

A much better connection has been established between workload and communications. A comprehensive study by Casali and Wierwille (1983) manipulated communication load during a simulated flight task; in addition to normal ATC instructions, the subjects were required to perform a call sign recognition task, with target call signs embedded in sets of extraneous call signs of varying difficulty. Of 16 workload measurement techniques employed, eight were sensitive to communication load manipulations. These techniques included both subjective ratings and objective measures. Hence, it is quite clear that communications load is a workload driver. However, the data reported by Casali and Wierwille (1983) does not allow for a reverse relationship to be established, that is, estimation of workload by analysis of the communication load. Several reasons prevent this: first, the article did not report any overall measures of communication load, such as number and durations of communications, and second, there were several other sources of workload present in the experiment, for example, piloting of the simulator. It should also be noted that a communication task is very different for a pilot and a controller. A pilot typically needs to respond to only a small fraction of messages transmitted on the frequency (i.e., only to those addressed to him or her), whereas the ratio of messages controllers receive and transmit is close to one (i.e., controllers talk to all aircraft on frequency).

Hurst and Rose (1978) replicated an earlier study that had indicated that peak traffic and the duration of radio communications were good predictors of behavioral response of air traffic controllers working in air route traffic control centers. This study included 3,110 observations made on radar sectors at the 13 major radar control rooms in the U. S. Duration of radio communications compared to behavioral ratings were made by expert-observer controllers showed that the former were good predictors of the latter.

A very strong relationship between controller workload and communications load was established in a study by Porterfield (1997). This study used ATC communications recorded from high-fidelity simulations and compared communication times to concurrently recorded subjective workload estimates (Air Traffic Workload Input Technique, ATWIT). The primary communications metric was average communication time per minute, calculated for 4-minute intervals to match ATWIT probes. A maximum coefficient of correlation of .88 indeed is very impressive, and the average communication time per minute also closely followed ATWIT ratings over a 15-minute period. However, the ATWIT ratings were generally very low, maximum ratings 3.5 on a scale from 0 to 7. At a workload rating 3.5 the communication load was 11 s per minute, or a proportion of .183.

Manning, Mills, Fox, Pfleiderer, and Mogilka (2002) analyzed 12 traffic samples from Kansas City Air Route Traffic Control Center (ZKC ARTCC). These traffic samples were viewed on SATORI (Rodgers & Duke, 1993) software, which recreated the traffic situations, by 16 ATC instructors who provided ATWIT workload estimates at 4-minute intervals. The samples were also processed by POWER software, which extracted a number of objective ATC taskload metrics from the data. Communications were quantified by the number of communication events and their durations, categorized by their content and speaker, as well as total communication times in 4-minute time epochs. The multitude of dependent variables was subjected to principal components analysis to reduce their number and like measures were combined to four taskload components. The results showed significant correlations between ATWIT ratings and total number and duration of communications ($r = .62$, $p < .01$), and individual communication durations ($r = .36$, $p < .05$), as well as number of instructional clearances ($r = .65$, $p < .01$). The activity component of taskload, which combined number of aircraft, number of simultaneously controlled aircraft, and radar controller data entries, was also correlated with total number and duration of communications ($r = .63$, $p < .01$), as well as with the number of frequency changes ($r = .36$, $p < .05$) and instructional clearances ($r = .52$, $p < .01$). Hence, it may be concluded that communication metrics may be a valid indicator of controller workload and taskload, although the r-values reported certainly leave other factors to be accounted for.

Availability of Data

Recent technological advances, particularly in area of digital technology, and the ATC modernization efforts potentially make available new sources for data as well as data collection and storage methods. An example of access to data from which various measures can be derived is the System Activity Recordings (SAR) that stores all flight and radar information in Air Route Traffic Control Centers (ARTCCs). These data can be further processed by two specific computer programs, the Data Analysis and Reduction Tool (DART) (Federal Aviation Administration [FAA], 1993) and the National Track Analysis Program (NTAP) (FAA, 1991), which produce a number of text-based output files. These files can be further analyzed by specialized computer programs, such as the Performance and Objective Workload Evaluation Research (POWER) (Mills, Manning, & Pfleiderer, 1999; Manning, Mills, Fox, & Pfleiderer, 2000). Currently, the POWER program derives over 40 separate measures that describe a variety of aspects of ATC.

Although a number of POWER measures have been shown to correlate with other sector complexity and workload measures, their relationship with controller performance is less clear (Manning, et al., 2000). On the other hand, ATC voice data has been shown to be a good indicator of controller workload (Hurst & Rose, 1978; Porterfierld, 1997; Manning et al., 2002), but they remain difficult to obtain and painstaking to analyze. If a valid relationship between certain complexity metrics and communication measures could be established, however, that would allow bypassing analysis of voice data in favor of mostly automatic data collection via POWER and similar tools.

Method

Data from three sectors from the Indianapolis air route traffic control center (ZID ARTCC) were selected for POWER analysis. The selection criterion for these sectors was simply that they should be very different from each other with unique characteristics in terms of traffic patterns and load. A senior supervisor from ZID chose the sectors based on these requirements and his expert judgment; the sectors were River (26, RIV) low-altitude sector, Dayton (88, DAY) high-altitude sector, and Wabash (99, WAB) super high-altitude sector. Two one-hour samples from each sector were obtained, one from busy and one from slow time of day.

Analysis of these data by POWER yielded many variables pertinent to sector complexity. In addition, voice data from the same samples were obtained and converted to wav files. These files were analyzed by SPWave program (SPWave is freeware and can be downloaded from http://www.itakura.nuee.nagoya-u.ac.jp/people/banno/spLibs/spwave/). This program allowed for visualization of the voice data as a spectrogram, and a zoom capability allowed for very accurate determination of transmission begin and end times. The data were coded (but not transcribed) and entered into an Excel spreadsheet. From the coded data a total of 53 variables were derived. These data were then compared to a number of complexity metrics that could be derived from the POWER output. Both voice and POWER data were examined in 10-minute epochs within the 1-hr samples.

Results

There were a total of 53 separate variables that were derived from the voice data. The results reported here, however, only pertain to those variables that either have been shown to correlate with controller workload and those that showed significant differences between the different ZID sectors. Furthermore, total number and duration of communications were highly correlated, as might be expected (R-squared = 0.854) and therefore only communication duration is discussed here.

Differences Between Sectors

As Porterfield (1997) and Manning et al. (2002) had discovered, communication time was a good predictor of workload (subjective ratings) and it was therefore of interest to examine whether the three ZID sectors differed from each other in this respect (see Fig. 1). An ANOVA on the proportion of controller communication time showed nearly significant (at a = .05) differences between sectors, $F(2, 29) = 2.90$, $p = .071$, and significant differences between busy and slow times, $F(1, 29) = 20.31$, $p < 0.001$. The interaction between sector and time (busy or slow) was not significant. These results, however, should be moderated by the small sample size, with only 6 data points (epochs) per condition.

Number of instructional clearances has also been associated with controller workload (Manning et al., 2002) and clear differences were found between the sample ZID sectors (Figure 2). An ANOVA showed significant differences between sectors, $F(2, 28) = 7.07$, $p < .0005$, and between times, $F(1, 28) = 19.09$, $p < .0005$. The interaction between sector and time was not significant, however.

Finally, we examined the number of frequency changes between sectors, as this variable has also been shown to correlate with workload. No statistically significant differences between sectors in the ZID sample were found, however, but time had a significant effect on the number of frequency changes, $F(1, 27) = 17.51$, $p < .0005$. This results is not surprising, as number of frequency changes strongly correlates with the number of aircraft in the sample, which clearly is the main difference between busy and slow times.

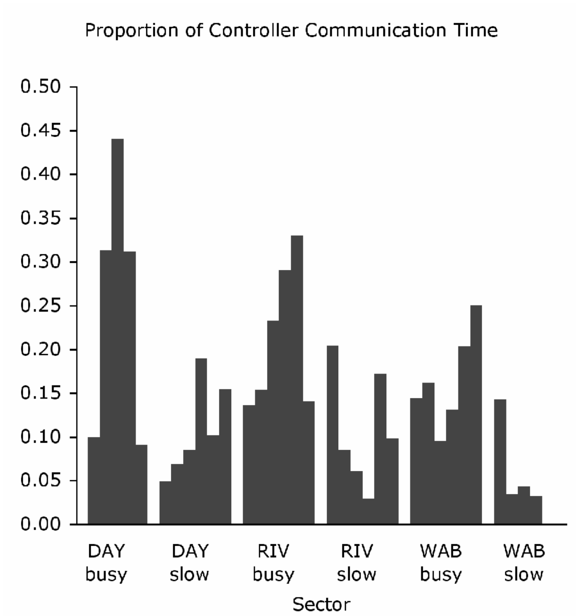


Figure 1. *Proportion of controller communication time in the six samples from ZID. Note that the maximum in DAY sector during busy time approaches 50%, meaning that the controller was speaking for almost half of the time during the 10-minute epoch. WAB had much lower communication load than the other two sectors.*

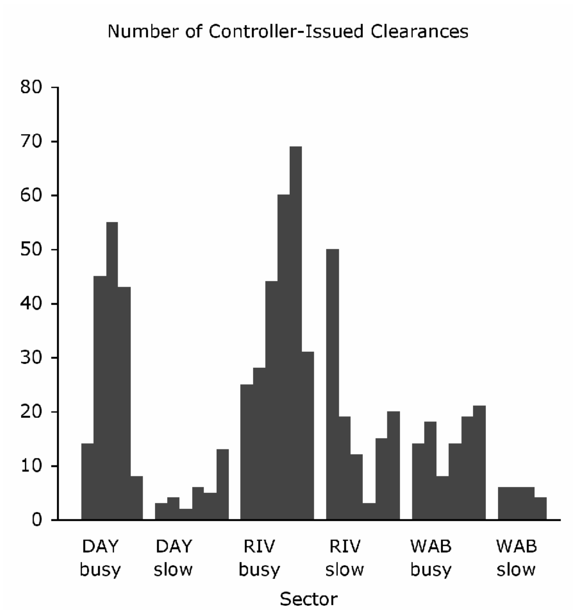


Figure 2. *Number of controller-issued clearances shows significant differences not only between times but also between sectors. RIV is clearly in a class of its own, as might be expected for a feeder sector.*

Aircraft Count and Communications Load

Given that communication time has been found to be a good predictor of workload (Manning et al., 2002; Porterfield, 1997), we examined correlations between the communication time recorded from the ZID voice data and POWER metrics from the same samples. Best correlation was found between the sum of three controller activity metrics (altitude changes + heading changes + number of handoffs) and controller communication time. The premise was that as aircraft altitude and heading changes currently necessitate a clearance, as does handoffs, they can be combined into an index that captures most of controller activity (Actvity Count). The results are depicted in Figure 3 below. A regression analysis showed a significant relationship between activity count and communication time, $F(1, 31) = 43.66$, $p < .0001$, R-squared = .5848.

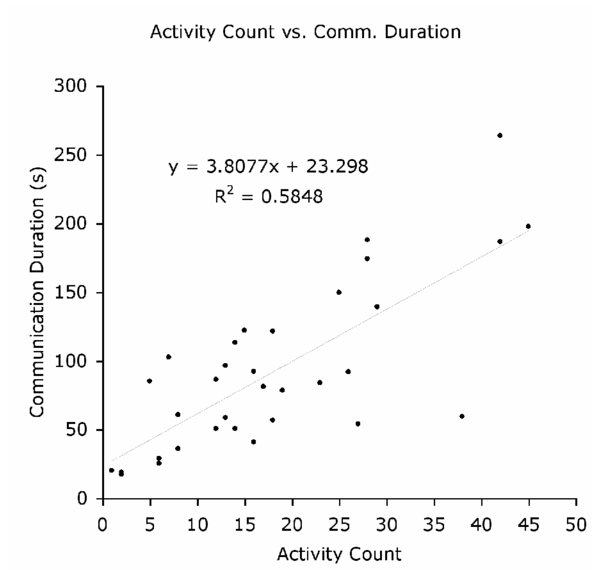


Figure 3. *Activity count, which is a sum of three POWER metrics (altitude changes + heading changes + number of handoffs) regressed against controller communication duration, which in turn has been shown to be a good predictor of workload.*

However, only slightly poorer results were obtained from regression of aircraft count and controller communication duration. A linear regression yielded a significant relationship between these variables, $F(1, 31) = 37.83$, $p < .0001$, R-squared = .5496 (Figure 4). Since aircraft count is much easier to obtain from data, this metric appears to suffice for an indicator of controller workload, inferred from communication duration.

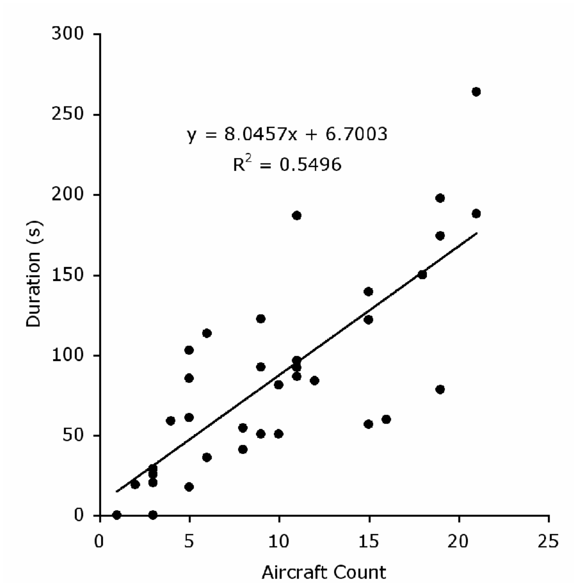


Figure 4. *Relationship between aircraft count and cumulative controller communication duration was positive and statistically significant. The correlation coefficient was only slightly poorer than that obtained from activity count.*

Discussion

Analysis of ATC voice data from three different sectors of ZID ARTCC revealed substantial differences between the sectors as well as between busy and slow times within the sectors. Given that communication duration and number of clearances issued have been shown to be workload drivers, we may conclude that the sample sectors can indeed be ranked in terms of workload imposed on the controller. In this respect, it appears that the high-altitude DAY and low-altitude RIV sectors were much more demanding than the superhigh-altitude WAB sector. Furthermore, it also appears that a simple metric of controller taskload, that is, aircraft count, correlated nearly as well with communication duration as did the more complex activity count, clearly favoring the use of the former as an indicator of controller workload. There are, however, several caveats that should be considered when assessing the validity of these conclusions.

First, a number of POWER metrics clearly differentiated between the sectors of different characteristics, revealing important factors that might affect controller taskload (e.g., maximum number of aircraft under controller's responsibility at any one time, proportion of aircraft changing altitude, handoff acceptance latency) that were not reflected in voice data. Equally important is to consider metrics that remained essentially invariant between sectors (e.g., number of aircraft), as these may reflect taskload factors that are independent from sector characteristics.

Second, although the POWER output included many parameters that were also part of the proposed airspace complexity and dynamic density measures as reviewed before, none of these metrics could be fully calculated for the sample sectors. Those complexity variables that were computed, that is, proportion of climbing and descending aircraft, average vertical distance between aircraft pairs, and aircraft density, did not show particularly strong correlations with communications measures. Finally, it must be acknowledged that the sample size in this study was quite small, with a maximum number of data points of 36 (3 sectors x 2 samples x 6 10-minute epochs) only marginally sufficient for regression analysis.

Nevertheless, this research may serve as an example for future validation efforts of various metrics of ATC complexity and controller taskload and workload. It should be kept in mind, however, that the aforementioned constructs are themselves complex and often involve multiple interactions, and hence simple measures may reveal only a partial picture of the situation.

Acknowledgments

This paper is based upon work supported by the Federal Aviation Administration under Collaborative Agreement No. 02-G-019. Dr. Carol Manning was the scientific/technical monitor. Any opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the FAA.

References

Casali, J. G., & Wierwille, W. W. (1983). A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors, 25*(6), 623-641.

Chatterji, G., & Sridhar, B. (2001). Measures of air traffic controller workload prediction. *First AIAA Aircraft Technology, Integration, and Operations Forum* (AIAA Paper 2001-5242). Los Angeles, CA, Oct. 16-18, 2001.

Federal Aviation Administration (1991). *Multiple Virtual Storage (MVS); Subprogram Design Document; National Track Analysis Program (NTAP).* (NASP-9247-PO2). Washington, DC: Author.

Federal Aviation Administration (1993). *Multiple Virtual Storage (MVS); User's Manual; Data*

Analysis and Reduction Tool (DART). (NASP-9114-H04). Washington, DC: Author.

Federal Aviation Administration (2000). *Dynamic density—A review of proposed variables*. NAS Advanced Concepts Branch, ACT-540, William J. Hughes Technical Center.

Hopkin, V. D. (1995). *Human factor in air traffic control*. Bristol, PA: Taylor & Francis.

Hurst, M. W., & Rose, R. M. (1978). Objective workload and behavioural response in airport radar control rooms. *Ergonomics*, *21*(7), 559-565.

Knecht, W., Smith, K., & Hancock, P. A. (1996). A dynamic conflict probe and index of collision risk. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 106-110). Santa Monica, CA: HFES.

Kopardekar, P., & Magyarits, S. (2002). Dynamic density: Measuring and predicting sector complexity. Paper presented at the 21st Digital Avionics Systems Conference, Irvine, CA.

Laudeman, I. V., Shelden, S. G., Branstrom, R., & Brasil, C. L. (1998). *Dynamic density: An air traffic management metric* (NASA-TM-1998-112226). Moffett Field, CA: NASA Ames research Center.

Manning, C. A., Mills, S. H., Fox, C. M., & Pfleiderer, E. (2000). Investigating the validity of performance and objective workload evaluation research (POWER). *Paper presented in the 3rd USA/Europe Air Traffic Management R & D Seminar*, Neaples, Italy, June 13-16, 2000.

Manning, C. A., Mills, S. H., Fox, C. M., & Pfleiderer, E. (2001). Investigating the validity of performance and objective workload evaluation research (POWER). DOT/FAA/AM-01/10. Washington, DC: Office of Aerospace Medicine.

Manning, C. A., Mills, S. H., Fox, C. M., Pfleiderer, E. M., & Mogilka, H. J. (2002). *Using air traffic control taskload measures and communication events to predict subjective workload*. (DOT/FAA/AM-02/4). Washington, DC: FAA Office of Aerospace Medicine.

Meshkati, N. (1988) Toward Development of a Cohesive Model of Workload.. In Hancock, P.A. and Meshkati, N. (Eds.) *Human Mental Workload*. Elsevier.

Mills, S. H., Manning, C., & Pfleiderer, E. M. (1999). Computing en route baseline measures with POWER. *Poster presented at the 10th International Symposium on Aviation Psychology*, Columbus, OH, May 3-6, 1999.

Pawlak, W. S., & Brinton C. R. (1996). Issues in free flight: Results from controller-in-the-loop simulations. *Air Traffic Control Association Conference*. Alexandria, VA: ATCA.

Pawlak, W. S., Brinton, C. R., Crouch, K., & Lancaster, K. M. (1996). A framework for the evaluation of air traffic control complexity. *AIAA, Guidance, Navigation and Control Conference* (San Diego, CA, July 29-31, 1996; AIAA Paper 96-3856). Reston, VA: AIAA.

Potrterfield, D. H. (1997). Evaluating controller communication time as a measure of workload. *The International Journal of Aviation Psychology*, *7*(2), 171-182.

Smith, K., Scallen, S. F., Knecht, W., & Hancock, P. A. (1998). An index of dynamic density. *Human Factors*, *40*(1), 69-78.

Sridhar, B., Sheth, K. S., & Grabbe, S. (1998). Airspace complexity and its application in air traffic management. *Proceedings of the 2nd USA/Europe Air Traffic Management R&D Seminar*.

Vidulich, M. A. (2003). Pilot actions and tasks: Selections, execution and control. In P. S. Tsang and M. A. Vidulich (Eds.), *Principles and practice of aviation psychology* (pp. 115-146). Mahwah, NJ: Lawrence Erlbaum.

Wickens, C. D., Mavor, M., & McGee, J. (1997). *Flight to the future: Human factors of air traffic control*. Washington, DC: National Academy of Science.

Wickens, C. D., Mavor, M., Parasuraman, R., & McGee, J. (1998). *The future of air traffic control*. Washington, DC: National Academy of Science

STRATIGIES FOR CONTOLLING CHECKLIST READING BEHAVIOR: A LITERATURE REVIEW

William G. Rantz
Western Michigan University
Kalamazoo, Michigan

One of the highest frequencies of errors recorded by recent Line Oriented Safety Audits (LOSA) is within the category of intentional non-compliance of which checklists use is included. These errors have led to serious lapses in risk management and many well- documented cases of aircraft accidents. This paper reviews the literature of both organizational behavior management and applied behavior analysis where checklist use is an independent variable. This report presents various methods and technologies from other settings which may prove useful in the flight-training environment. Also included is a proposed study that will be conducted at a major flight training facility using undergraduate participants involved in checklist use while undergoing instrument flight training. This study applies various treatments to the participants to measure the effectiveness of checklist reading behavior and performance. Measures examine both, short term and long term effects of treatment, as well as any generalization of checklist reading performance to more advanced training environments.

Checking Checklist Performance

Upon reviewing the checklist literature it becomes apparent that much has been documented regarding checklist design, checklist importance (Degani and Weiner 1990, FAA, 1995), and accidents and incidents resulting from the misuse of checklists (NTSB, 1969, 1975, 1982, 1988a, 1988b, 1989, 1990, 1997). Yet there seems to be a lack of studies regarding effective technologies that will improve a crew's use of checklists in flight operations. Using the Line Oriented Safety Audit (LOSA), Helmreich reported over 50% of all in flight errors were intentional non-compliance. Non-compliance errors were, "conscious violations of standard operating procedures (SOP) or regulations. Examples include omitting required briefings or checklists" (Helmreich 2000).

Checklist performance may vary widely between operators. (Diez, Boehm-Davis, and Holt, 2003). Two methods of checklist design are the challenge-do-verify (CDV) and the do-verify (DV). These methods can be paired with text/paper, mechanical, or electronic aids to ensure the crew is not relying only on memory (FAA 1995). Degani and Weiner (1990) identified similar checklist methods using challenge and response, memory-guided, and short-cutting or "chunking" the checklists. Chunking comprised calling a limited number of challenge items from the checklist, then checking those tasks by reading from the checklist.

Examining the checklist accident data, evidence would suggest that proper checklist use is vital in performing sequential tasks at the appropriate time in a potentially highly distractive environment. One question should be: How can we train and bring checklist behavior to a level of consistency that SOPs require? Another question might ask; Is there other variables within the training environment which may be paired with checklist reading behavior which might increase or maintain the frequency of checklist use?

Diez, et al (2003) identified significant differences in compliance with crews of one airline using a memory-based checklists. Verbal annunciation of checklist items was a required SOP for the airline of interest. Of the expected 100% compliance, the crew only vocalized approximately 78% of the checklist items. These differences were observed across aircraft and between checklists. Searching for checklist studies, such as the one above, where the checklist is the independent variable in aviation literature, is limited.

Experiments in other disciplines, which use checklists as a component of the independent variable, may provide important clues relating to any changes in performance resulting from manipulations of those variables. Through experimental design, performance in checklist use can be measured. Therefore the resulting intervention, and subsequent change in performance, may provide application in a flight training setting.

Literature Review

The purpose of current review is to (a) examine the literature for checklist use in organizational performance change, (b) determine those studies where checklist use was most effective, (c) discuss effectiveness of interventions where checklists were paired with other variables, and finally recommendations are made for future research in the area of checklist use in flight training.

Method

Articles were reviewed from the *Journal of Organizational Behavior Management* between the years 1977 and 2003 and the *Journal for Applied Behavior Analysis* between the years 1968 and 2003. Particular keywords, listed below, were used to search in the PsycInfo 1887 database for relevant articles.

Inclusion and Exclusion Criteria

Inclusion criteria were the keywords, (a) checklist, (b) job aid, (c) task analysis, (d) task clarification, (e) prompt, (f) task performance, (g) self-monitoring, and, (h) task list. Both journal titles were used concurrently with each keyword to generate search results.

Only studies in organizational contexts were used with the intervention being applied to real-world tasks. If it could not be determined that the participants were employed in the target organizational environment and given some form of a written checklist to use during performance of the task, the article was excluded. Articles in which target participants were mentally impaired or in school environments were excluded. This provided inclusion of studies where responses were prompted, which resulted in, a product, service or some other measurable accomplishment.

Under these criteria nine articles were discovered for review. One article of the nine contained three separate studies, which increased the total review to eleven studies of checklist use as the independent variable. Three other articles, from other journal sources were discovered, due to their contribution to the subject of checklist use, they will be discussed. These articles are not included in the final review and will only provide supporting discussion.

Interobserver Agreement

A trained rater independently coded each article in the review. Inter-observer agreement was calculated using this formula: (Disagreements/Agreements + Disagreements) x 100. Reliability was 100%.

Results

The literature review revealed multiple settings and tasks where checklists were used as a stand alone component or in combination with other independent variables. Two articles used a manufacturing environment as a setting. The first, Moses, Stahelski, and Knapp, 2000, used control charts and a check sheet process as a strategy to control reducing the size of metal tubes. The second study used posted written set-up procedures to reduce set-up times on a die extruder machine, Wikoff, Rowan, and Poling, 1990. Four studies involved the hotel industry. Three of the four studies were contained in one article by Anderson, Crowell, Sponsel, Clarke, and Brence, 1983. Checklists were used with various cleaning, room preparing, and support tasks. The forth article by LaFleur and Hyten, 1995 used checklist for preparing banquet rooms for hotel customers. The only study to use checklists in an office setting was by, Bacon, Fulton, and Malott, 1982. This study examined the tasks of record keeping, grading, lesson completion, and system maintenance. One study by Porterfield, Evans, Blunden, 1985, demonstrated checklist use in the form of a distributed leaflet, to improve performance of health care workers with developely disabled individuals. Another interesting study used a student managed bar to conduct an intervention of checklist posting and feedback to improve cleaning tasks, Anderson, Crowell, Hantula, and Siroky, 1988. For a down home approach, Altus, Welsh, and Miller, 1991, used checklists in a student housing cooperative to improve task performance in a domestic setting. Customer service tasks were improved by including a posted task list for bank tellers in a study by Crowell, Anderson, Abel, and Sergio, 1988. For a complete comparison of the review studies, Table 1 summarizes the findings of the review.

Review Analysis

Checklist medium. All studies used a written checklist format. Some examples were provided within the studies to compare the compositions (Anderson et al. 1988; LaFleur et al. 1995; Moses et al. 2000). Eighty two percent of the studies described the checklist as some type of written document that could be posted or carried by the participant. As previously mentioned in the Crowell, Anderson, Abel, and Sergio (1988) study, the checklist could be as simple as a memo listing behavior prompts or in the Porterfield, Evans, and Bluden (1977) article which described the staff roles as a leaflet to be used at the discretion of the participant. Other studies demanded more involvement of the participants with the checklist, such as office task work by Bacon, Fulton, and Malott, (1982), which required participants to mark items on the checklist as the task progressed.

Checklist pairings. All the studies in this review, except one, used checklists as one component part of an intervention. Only the Bacon, Fulton, and Malott

(1982) study used the checklist alone as an independent variable. The remaining ten studies paired the checklist with various forms of feedback, tokens, goal setting, or a punisher in the form of a fine. With all studies, the dependent variable measurably changed in the desired direction. However, one drawback is that a component analysis was not conducted in any of the studies with checklist pairing. Therefore it is difficult to conclude that the results produced by the interventions were due to paired checklist use or checklist use alone. It is assumed observed effects are from the pairing of checklists in combination with various forms of feedback. Many times it was difficult to determine from the study how consistently the checklist was used and therefore was it or was it not consistently paired during the intervention.

One manufacturing study may have paired a checklist with an attribute control chart for feedback (Moses et al. 2000). This particular chart tracks specific control limit events using statistical process control. It is not clear from the study if the participants who used the checklist to sequence the inspection process actually viewed the attribute control charts. This leaves a potential confound in the study with regard to whether pairing intervention components actually occurred.

The second manufacturing study paired the checklist of set up procedures with two feedback methods, observation audits and video feedback (Wikoff et al.1990). Prior to the intervention, the experimenters conducted a task analysis and listed sequential steps in the set-up procedure, which would yield optimum performance. Copies of the written set-up procedures were given to each participant and one copy was posted on each machine. It is not conclusive from this study that the participants actually used the set-up procedures checklist each time they set up their machine. At least once each week, for four months, each participant was video taped and feedback provided regarding performance of set up times. After the four-month time period a trained supervisor conducted an operational audit at least once each week, for three months. Verbal feedback was provided by supervisor regarding the participant's performance. Set up time did decrease during the intervention yet without controlling for consistent checklist use or conducting a component analysis, it is difficult to judge the effectiveness of the checklist or with the checklist in either feedback combination.

The six service studies used a variety of feedback methods paired with checklists. Three of those studies conducted by Anderson, Crowell, Sponsel, Clarke, and Brence (1983) used room cleaning, housemen, and doormen checklists paired with weekly posted charts of completed checklist items. This pairing was followed by a period of several weeks into the intervention with the same checklists paired with the same-posted chart and adding tokens awarded for criterion or better performances. Again, it is difficult to tease apart the effectiveness of each component in these interventions. There was a desirable directional change in many of the performances, however no definitive conclusions can be drawn as to which individual component or combinations of components may yield similar results.

The study by Porterfield, Evans, Blunden, 1985 investigated health care workers. During this four-phase experiment, the experimenters added the leaflet checklist and the daily observable feedback to the participants during the same phase. This procedure missed an opportunity to measure each component separately within the study. From the pairing of both the leaflet and the vocal feedback, it is evident that desirable behavior increased as illustrated in the study results.

The study of banquet set-up tasks was conducted by LaFleur and Hyten 1995. During this study task checklists were explained in detail to the staff. This study also used the checklist as a response sheet to record when each task was complete by signing their name beside each completed task. The checklists were later collected by the participants' supervisor. Checklist use was simultaneously paired with publicly displayed, daily, setup completion percentages graphs, goal setting, and monetary bonuses. Results indicate that setup completion percentages increased from between 40 percent variable to 100 percent to nearly 100 percent consistently. This provides supporting evidence to the effectiveness of the total treatment package. However conclusions can only be made that the checklist usage in food setup tasks may be effective when used with some type of supporting reinforcing intervention.

The student operated bar paired task clarification with posted checklists in an experimental design that isolated this portion of the intervention from the feedback portion (Anderson et al. 1988). Feedback was provided in the form of publicly displayed line graphs. This study does show evidence that antecedent prompting of task clarification and posted checklists can make an immediate change in desired behaviors. The study states a sample of an unscored checklist was posted for continuous viewing. The study indicated that during the portion of antecedent-like treatment, behaviors increase modestly. Behaviors increased again after the line graph feedback treatment was added to the intervention.

This study provides some evidence of behavior change prior to the addition of feedback. However the treatment protocol is vague regarding the consistent use or viewing of the checklist by the participants during the first phase of the intervention. Yet the evidence suggests that some type of effect occurred as a result using paired task clarification and posted checklists.

A similar intervention was used in the bank teller study by Crowell et al. (1988). In this study task clarification was explained followed by a "clarification" memo given to all participants explaining the behavioral categories and description of the point system for scoring performance. The study also paired the task clarification with a graph of mean transaction quality points, verbal feedback, and praise. The authors report, "performance change produced by clarification emerged quickly and remained relatively consistent throughout the phase" (p. 69). The study also reported, "present effects of task clarification are noteworthy because they are consistent with prior evidence showing that knowledge of task relevant behaviors can facilitate work performance, even in the absence of explicit feedback" (p. 70). This statement suggests that the participants gained knowledge of the tasks to consistently perform to a level higher than baseline. The duration of the task clarification phase intervention was 35 days. During this time the task knowledge was either drawn from memory of the initial task clarification briefing or from frequent review of the task clarification memo. This study does illustrate changes in desired behaviors during each phase of the experiment. Yet the study is unclear with regard to the experimental protocol concerning use of checklists to prompt the complex tasks of social interaction of bank tellers.

The Altus et al. (1991) study examines mean percentage of task completion of household duties by following a written checklist of those tasks. The checklist was paired with tokens for adequately completed tasks as well as fines for a specific level of uncompleted tasks. This is the only study in the review, which used an obvious punisher. Fines were assessed for behavior other than on task, paired with a checklist during the intervention. Behavior changed in a desirable direction using the paired intervention of written checklists and tokens and fines. The researchers concluded that the participants managed the tasks very effectively after the introduction of the checklists. The study suggests the resulting increase in task completion and the decrease in fines and complaints support the notion that the pairing of written checklists with rewards and punishers may have merit. However the study did not go into detail regarding potential confounds of using a punishment technique in the study. Counter control issues related to punishment or emotional bursts were never discussed.

The study, which isolated the checklist as an intervention, was the Bacon, Fulton, and Malott (1982) study. However this study did use three specific conditions from a study conducted by Brethower (1970), which required participants to understand and interact with the checklists. Except for the LaFleur and Hyten (1995) study, it was not evident that any other studies used all of these three elements as did Bacon et al. 1982. The first required element was the need to view the checklist daily (i.e., this assumes the required frequency of the task is daily). In doing so, the checklist requires evidence of completing the tasks. The second required element is recording, on the checklist, a specific time of task completion or amount of work that has been completed. The last requirement, to maintain checklist usage, is periodic review of the checklist by the participant's supervisor. It was clear from the study that the participants would not receive any punishing consequences resulting from an incomplete checklist. Each week the participants would receive a new checklist without any mention from the researcher of the previous checklist. Therefore no feedback was given to the participants. The results of the study seem to demonstrate a desired change in task completed behaviors by using the checklists alone. The study also discussed the potential for reinforcing effects from using the checklist alone. Some examples of contingencies for continued checklist use might be the result of rule-governed behavior. The potential punishing consequences of the supervisor reviewing an incomplete checklist may support continued checklist use. The viewing of the checklist could also provide a reinforcing effect for actually working on the tasks by seeing the task checked off the list.

As with most of the studies reviewed, the pairing of checklist use was done without a component analysis to determine if on task behavior would result from checklist use alone. Perhaps many of the effects demonstrated in these studies are under the influence of multiple contingencies that are yet to be isolated and control of those contingencies demonstrated.

Discussion

The review articles claim that checklists have been used in capacities to prompt specific behavioral performance. Many of these studies use interventions with multiple components, which are paired with checklists, to produce behavior change. These paired components ranged from task clarification of the environment and task knowledge, to incentives, graphic feedback, and punishment. Gilbert (1978) supported the notion that for workers to perform well they require environmental information, knowledge, and response supports. As one of the intervention components, checklists may provide response supports, which increase the efficacy of complex task completion.

Most checklists within the review articles were developed from the need to document specific behavioral response chains. This effort requires a detailed job task analysis of the desired knowledge, skills, and abilities of the specific task (Gatewood, and Field, 1998). One approach to developing a checklist is to conduct a thorough job task analysis, which should identify the present skills, and abilities that demonstrate a particular level of performance. Using information from the job task analysis, a checklist can be constructed that will maximize performance sequences and provide the worker task clarification (Anderson et al., 1988, Degani and Weiner, 1990).

Checklists can be used to prompt specific behavior to occur. Prompting of behaviors As such, training, job aids, task clarifications, and checklists can be antecedents for behavior. While many of the studies demonstrated behavioral changes due to the use of checklists, only the Bacon et al. study in the literature review used checklists alone to effect behavior change. All other studies eventually linked checklists with some form of reinforcer or punisher as a total intervention package. This leaves to question the behavioral function of the checklist and what contingencies may support continued checklist use

One article not included in this review is by Shier, L., Rae, C., and Austin, J. (2003) which used five checklists to improve performance in a grocery store environment. The checklist were signed and returned to the researchers who then graphed the results and used that data for public posting. Again this intervention used task clarification, checklists, and feedback to demonstrate performance change. The most unique element of this study was the authors attempt to conduct an organizational functional assessment (Austin, Carr, and Agnew, 1999) to determine cause of the poor performance.

Due to the lack of empirical research on specific, stand alone, checklist systems, the exact behavioral functions of checklist use remain for future research.

The issue of checklist use and distraction, in a particular setting, is not addressed to any extent in most of the studies. No study addressed a treatment protocol for checklist handling. Assumptions were made that participants would use the checklist as needed perhaps depending on the strength of other contingencies for not using the checklist (Lafleur et al. 1995, Bacon et al. 1982, Moses et al. 2000, Anderson et al. 1988). No study examined the consistency of the environment where the checklist was used. Future research should examine checklist use in changing environments where the potential for error in checklist reading could be high and the penalty for error would be costly such as flight training

.

Task clarification and checklists seem to produce a rapid change in behavior immediately and consistently after introduction (Bacon et al. 1982 and Crowell et al. 1988). As an effective, inexpensive method of improving and maintaining performance why is there not more dedicated research focused on the use of checklists alone? Only the Bacon et al. 1982 study tried to examine the use of checklists without pairing them with other interventions. Future research should investigate refinements of checklist construction. Studies could examine how the checklist interacts with the users. There should be a validation process to determine if some type of supervisory contingency must be created to maintain checklist use or can an interactive contingency be created with the checklist and the user alone.

Conclusion

This review compared the use of checklists in the literature to determine areas of commonality in the field of applied organizational studies. It is evident that service tasks have been the focus of many researchers. These tasks and settings present potential for checklist interventions by the nature of the somewhat transient populations, repetitive nature of the tasks, pay scale of the jobs, and the level of detail required completing the tasks. All of the studies were in applied setting which present unique challenges regarding experimental control. However each study did demonstrate some type of behavioral change through the use of an intervention. Unfortunately most of the interventions used some type of checklist pairing procedure without examining checklist effects independently prior to pairing with another component.

It is very likely that using similar checklist pairing procedures may result in performance changes in flight training performance. Checklist strategies have long been paired with other interventions. Perhaps it is time to examine in finer detail the changing technologies in which checklist use can be created, monitored, and refined. Generalizing from the performance successes reported in the literature review, future checklist research in flight training will isolate and test the antecedent, consequence, and motivating operations associated with checklist use in stable, threatening, and changing environments. Dependent variables will consist of observable behaviors in checklist reading i.e. fluency, frequency, latency, ratios of items performed per segment and ratios of items performed correctly over time. The manipulation of the independent checklist variable will consist of pairing checklist use with and without graphic feedback of the dependent variables. The research protocol will use PC-based flight trails while conducting an instrument approach. It would be of great benefit to confirm the reliability and validity of checklist pairings with various settings and tasks.

Using checklists has a long history in many settings and for many tasks. It is time to look closer at the checklist and determine if there isn't more to using the checklist than already exists in the literature.

References

Altus, D., Welsh, T., and Miller, K. (1991). A technology for program maintenance: Programming key researcher behaviors in a student housing cooperative. *Journal of Applied Behavior Analysis*, *24*(4), 667-675.

Anderson, C. D., Crowell, C. R., Hantula, D. A., and Siroky, L. M. (1988). Task clarification and individual performance posting for improving cleaning in a student-managed university bar. *Journal of Organizational Behavior Management*, *9*(2), 73-90.

Anderson, C., Crowell, C., Sponsel, S., and Brence, J. (1982). Behavior management in the public accommodations industry: A three-project demonstration. *Journal of Organizational Behavior Management, 4*(2), 33-66.

Bacon, D. L., Fulton, B. J., and Malott, R. W. (1982). Improving staff performance through the use of task checklists. *Journal of Organizational Behavior Management, 4*(4), 17-25.

Brethower, D. M. (1970). The *classroom as a self-monitoring system.* Unpublished dissertation, University of Michigan, 1970.

Crowell, C., Anderson, C., Abel, D., and Sergio, J. (1988). Task clarification, performance feedback, and social praise: Procedures for improving the customer service of bank tellers. *Journal of Applied Behavior Analysis*, *21*, 65-71.

Degani, A., and Wiener, E. L. (1990). Human factors of flight-deck checklist: The normal checklist. (No, NASA CR-177549). Ames Research Center, Moffett Field, CA: NASA.

Diez, M., Boehm-Davis, D., Holt, R. (2003). Checklist performance of the commercial flight deck. *Proceeding of the Twelfth International Symposium on Aviation Psychology*. (323-328). The Ohio State University.

FAA. (1995). Human performance considerations in the use and design of aircraft checklists. Washington DC: Federal Aviation Administration.

Gatewood, R., and Field, H. (1998). *Human resource selection* (4th ed.). Fort Wayne, IN: The Dryden Press.

Gilbert, T. (1978). *Human competence: Engineering worthy performance.* Washington D.C.: International Society for Performance Improvement.

Helmreich, R.L. (2000). Culture and error in Space: Implications from analog environments. *Aviation, Space, and Environmental Medicine*, 71(9-11), 133-139.

LaFleur, T., and Hyten, C. (1995). Improving the quality of hotel banquet staff performance. *Journal of Organizational Behavior Management*, *15*(1-2), 69-93.

Moses, T., Stahelski, A., and Knapp, G. (2000). Effects of attribute control charts on organizational performance. *Journal of Organizational Behavior Management*, *20*(1), 69-90.

NTSB. (1969). *Aircraft accident report: Pan American World Airways, Inc., Boeing 707-321C, N799PA, Elmendorf Air Force Base, Anchorage, Alaska, December 26, 1968* (No. NTSB/AAR-69/08). Washington, DC: National Transportation Safety Board.

NTSB. (1975). *Aircraft accident report: Northwest Airlines, Inc., Boeing 727-25, 264US, Near Thiells, New York, December 1, 1974* (No.NTSB/AAR-75/13). Washington, DC: National Transportation Safety Board.

NTSB. (1982). *Aircraft accident report: Air Florida, Inc., Boeing 737-222, N62AF, Collision with 14th Street Bridge, Near Washington National Airport, Washington, DC., January 13, 1982* (No.NTSB/AAR-82/08). Washington, DC: National Transportation Safety Board.

NTSB. (1988a). *Aircraft accident report: BAe-3101 N331CY. New Orleans International Airport. Kenner, Louisiana* (No. NTSB/AAR-88/06).

Washington, DC: National Transportation Safety Board.
NTSB. (1988b). *Aircraft accident report: Northwest Airlines, Inc., McDonnell Douglas DC-9-82, N312RC, Detroit Metropolitan Wayne County Airport, Romulus, Michigan, August 16, 1987* (No.NTSB/AAR-88/05). Washington, DC: National Transportation Safety Board.
NTSB. (1989). *Aircraft accident report: Boeing 727-232, N473DA. Dallas-Fort Worth International Airport, Texas* (No. NTSB/AAR-89/04). Washington, DC: National Transportation Safety Board.
NTSB. (1990). *Aircraft accident report: USAir Inc., Boeing 737-400, LaGuardia Airport, Flushing, New York, September 20, 1989* (No. NTSB/AAR-90/03). Washington, DC: National Transportation Safety Board.
NTSB. (1997). *Aircraft accident report: Wheels up landing, Continental Airlines Flight 1943, Douglas DC-9 N10556, Houston, Texas, February 19, 1996* (No. NTSB/AAR-97/01). Washington, DC: National Transportation Safety Board.
Porterfield, J., Evans, G., and Bluden, R. (1977). Involving families and staff in service improvement. *Journal of Organizational Behavior Management*, *7*(2), 117-133.
Shier, L., Rae, C., and Austin, J. (2003). Using task clarification, checklists and performance feedback to increase tasks contributing to the appearance of a grocery store. *Performance Improvement Quarterly*, *16*(2), 26-40.
Wittkopp, C., Rowan, J., and Poling, A. (1990). Use of a feedback package to reduce machine set-up time in a manufacturing setting. *Journal of Organizational Behavior Management*, *11*(2), 7-22.

HUMAN FACTORS AND HUMAN RESOURCES DEVELOPMENTS FOR PAN-EUROPEAN IMPLEMENTATION: ACHIEVEMENTS IN THE EUROPEAN ATM PROGRAMME

Hermann Rathje
European Organisation for the Safety of Air Navigation
(EUROCONTROL)
Human Factors Management Business Division
Brussels, BELGIUM

The European Air Traffic Management (EATM) Human Resources Programme aimed to deliver harmonised tools and a body of knowledge for the management of human issues in ATM in the three areas training, manpower (human resources management) and human factors. Products are available as Guidelines, as Technical Reference Material or Reports or as Methods and Tools for direct application. The four year work programme consisted of specific developments, testing and validation of these products available since end 2003 for implementation in the 41 European Civil Aviation Conference (ECAC) States. The implementation and use is not mandatory but products are applied increasingly with the active involvement of stakeholders from Air Navigation Service Providers (ANSPs), military and regulatory authorities and professional associations. Four different products are presented in this paper: Team Resource Management (TRM), Human Error Analysis (HERA), Critical Incident Stress Management (CISM) and the First European Air traffic controller Selection Test package (FEAST).

European ATM Framework

EUROCONTROL[1], the European Organisation for the Safety of Air Navigation is involved in the development of a seamless, pan-European Air Traffic Management (ATM) system to cope with the growth in air traffic, while maintaining a high level of safety, reducing costs and respecting the environment.

The EUROCONTROL Agency is tasked by the ECAC Transport Ministers with defining a common European vision and strategy and coordinating its implementation.

Under the performance enhancement programme for European ATM (EATM), the EUROCONTROL Agency produces standards and guidelines and common products / systems and tools and provides guidance and assistance to its Member States in the implementation thereof.

The European Commission, the executive body of the European Union[2], is now progressing with the creation of a Single European Sky that aims to enhance current safety standards and support commercial and economic growth through more efficient airspace design following operational needs rather than national frontiers, to generally optimise capacity and ensure interoperability of the ATM systems across Europe.

The Human Resources Programme (HRS)

The EATM programme consists of a wide portfolio of programmes, services and support activities and includes Human Factors, Manpower, Human Resources Management and Training activities.

The objective of the latter programme is "to ensure human involvement and commitment to support the change to future ATM so that operational, technical and support staff can operate effectively, efficiently and safely within their capabilities and obtain challenge and job satisfaction."

ATM systems are expected to remain human-centred for the foreseeable future, and people will play a key role in achieving system safety and capacity enhancements.

People are therefore an essential element in the ability to deliver ATM services, and their co-operation and involvement in developing and effecting change is essential.

It is of high importance that all human performance and training issues are sufficiently addressed and managed as early as possible, in order to ensure new technologies and operational procedures. This will enable stakeholders to proactively plan and manage their medium and long-term goals for the management of human issues in European ATM.

The aim of the Human Resources Programme is to offer, through the production of guidance material, methods and tools, a harmonised and integrated approach to:

[1] Numbered 34 Member States (in Dec 2002): Albania, Austria, Belgium, Bosnia, Bulgaria, Croatia, Cyprus, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Luxembourg, the former Yugoslav Republic of Macedonia, Malta, Moldova, Monaco, the Netherlands, Norway, Poland, Portugal, Romania, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine and the United Kingdom.

[2] The European Union (EU) comprises of 25 Member States: Austria, Belgium, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden and the United Kingdom.

- Manage human performance proactively and to ensure the timely availability of suitable operational staff through tools and methods for manpower planning, recruitment and selection, training and staff development;
- Enhance safety in day-to-day ATM operations through human factors products and tools for Team Resources Management and Critical Incident Stress Management and through tools that support integration of human factors into the life cycle of ATM systems;
- Progress with ATS staff training towards common standards in line with the regulatory requirements for controller licences and the changing ATC systems.

Next Steps in Human Factors Developments

The cultural, social, human factors and human resources aspects related to the intended reorganisation of the current European ATM system in line with the European Union Single European Sky initiative will be even more important to be appropriately addressed in the future. It will require developing new approaches and tools to

- effectively deal with the cultural, organisational and individual change and transition issues involved;
- fully integrate human factors into safety management systems and safety culture;
- provide common European training standards and tools in line with regulatory requirements

These work areas are now addressed in am proposed new Human Performance and Training Enhancement Programme to start later this year.

Human Resources Programme Product Development, Testing and Validation

During the years 2000 – 2003 the human resources work programme proceeded in parallel in 17 training, manpower and human resources management and human factors projects[3]. Four projects are reported in more depth in this paper:

- Human Error Analysis (HERA)
- Team Resource Management (TRM)
- Critical Incident Stress Management (CISM)[4]
- First European Air traffic controller Selection Test package (FEAST).

The general approach followed in the Human Resources Programme was to develop the projects in close consultation with stakeholders – Air Navigation Service Provider organisations (ANSPs), military and civil authorities and professional groups as well as with external partners in a coordinated fashion to ensure a broad representation of stakeholder requirements and needs. This also helps to later facilitate practical implementation and customisation of the products.

The needs, benefits and the feasibility for development were established in early feasibility studies or business cases that took stock of what was already available in the area of work. Sound feasibility reports, cost-benefit considerations and development options aim to market the intended work, clarify deliverables and justify the work in general with the aim to gain stakeholder commitment and support.

Prototypes of the tools were tested in practice and validated against established criteria. This early feedback and data is used to refine the work programme and direct further work. In all cases in which this was applicable beta-versions of the final products were tested in different national and local environment, representing a good cover of the cultural operational and administrative working environment. The outcome of the trials was validated in terms of <u>content</u> ('Does what is developed represent the subject area or behaviour it intends to represent?') and / or (concurrent) <u>criterion</u> validity ('Do results from using the product correlate with a relevant external criterion?'). Lessons learned from the 'live' trials were used in final updates of the products and reported.

Product Implementation

The ultimate purpose of the new products or is to provide valuable, scientifically sound, harmonised and cost effective options to ANSPs for use in training, human resources management, human factors and safety. Implementation is done in a coordinated fashion across ECAC States based on an agreed common action plan and target dates. There is an annual follow-up on the implementation actions in regard to all products across States that participate in the European ATM Programme. Air Navigation Service Providers, that decide to implement a product or tool are given professional assistance and support by the training, human resources or human factors experts from the development unit. Implementation support consists of

- product user training and certification;
- technical advise and support;
- planning and customisation support;
- functional helpdesk;
- administration and application enhancement and improvements;
- sponsoring and facilitating user Group meetings.

---

[3] Most deliverables are to be found on the following website: www.eurocontrol.int/humanfactors

[4] CISM was a project outside the HRS Programme as part of the Human Factors Domain activities.

The follow-up and user support activities are of great importance to ensure that products are used consistently and for the correct purposes as intended. The huge cultural and administrative and / or operational diversity across European States, the differences in organisational size and 'maturity' as well as the availability of local expertise require a sensible and sometimes sensitive approach in implementation support. There is no 'One size fits all'.

HERA – A New Technique for Human Error Analysis in ATM

Human error is a key contributor to risk for incidents (and accidents) in ATM and finding and mitigating if not avoiding the root causes or causal factors of human error is hence an important aspect in safety research and safety management. The importance of finding the root causes for human error in ATM is highlighted by recent statistics showing, in the US for example, an increase of human errors by more than 25% during the time period 1998 - 2003.

The research and development of a common approach and tools for human error analysis was jointly done by Eurocontrol and FAA in the period 1999 – 2003. The outcome was a new technique for the classification and assessment of the causal factors for human error, called JANUS[5] a technique originally used to retrospectively analyse ATM incident reports during investigations only. The idea was to use the taxonomy also to diagnose the potential for human errors prospectively, for example to assess potential human errors in the design of (future) systems. The collaborative Eurocontrol – FAA development process proceeded in four stages: planning, development of JANUS, field testing and validation.

Development proceeded on parallel tracks in Europe and in US: Eurocontrol developed the human error taxonomy following the HERA approach using ATM task and behavioural requirements and looking at the cognitive processes that lead or could lead to an incident. The partner organisation FAA used their Human Factors Analysis and Classification System (HFACS) approach that captures the conceptual breadth and depth of the system with the individual actions, along preconditions, supervision and organisational influences. Both approaches are hence complementary to some extent and had a good track record in previous validation studies. JANUS integrates both approaches in a common taxonomy of human errors and causal (cognitive) human factors.

Table 1 provides the JANUS taxonomy categories and examples.

Table 1. *JANUS Taxonomy*

Error Category	Subcategory / Examples
Error Type – How error was manifested	Action omitted, right action but wrong object
Error Detail – Which cognitive domain failed	Perception, Memory, Planning and Decision making, Response
Error Mechanism – What happened?	Late detection of information
Information Processing - Why did it happen?	'Tunneling', forgot to monitor
Contextual Conditions	Pre-conditions: airspace, teamwork, supervision, organisational factors

The JANUS technique itself consists of a series of flow diagrams (paper based) used in interview sessions with specially trained users (i.e. investigation experts). Investigators or researchers are systematically led through a series of questions one at the time. This reduces the occurrence of user bias and prevents jumping to conclusions.

JANUS Testing and Validation

Beta testing of the common taxonomy JANUS took place in seven European member States analysing a total of 60 incidents (done by Eurocontrol). The FAA independently applied JANUS to 79 incidents from 12 US facilities.

The findings from both parallel studies were analysed with a view towards five 'validity' questions:

- Does JANUS work?
- How well does it work?
- Is it better than current methods?
- Is it ready for implementation?
- Do results improve safety management?

The findings (objective / subjective reports) indicate that the technique works, is (moderately) consistent in identifying causal factors and helps to improve the investigation of human factors related incidents. JANUS identified on average 13 causal factors per incident compared to 2 factors from current methods used. The findings also showed that JANUS broadens the current scope of investigation substantially. It prompts investigators to causal factors in a given context situation in which a human error occurred.

[5] Janus is the name of a mythological figure (the Roam god of gates and doors) who, with his two faces looking in opposite direction. Janus represents the beginning and end, the past and the future and the transition from a less developed towards a more advanced stage of cultural live.

In summary: JANUS is more sensitive, is useful, comprehensive and practical than current available methods.

The benefits demonstrate the value of a joint undertaking and using a wider scope of expertise and experience. This has led to more consistent, sensitive and comprehensive approach in analysing and subsequently preventing or mitigating human errors in ATM based on common terminology. An important step towards international standardisation in this field has been achieved.

In the European context JANUS is seen as a means of complying with the European Safety Regulatory Requirements (ESARR). The next planned step is to have the JANUS taxonomy included in the European Co-ordination Centre for Aviation Incident Reporting System (ECCAIRS).

Team Resource Management in European ATC: A Ten-year look-back

Airlines have it since more than 25 years now and apply it around the world: programmes to promote teamwork practices in Crew Resource Management (CRM). Wiener etc al. (1993) noticed the lack of it in ATC.

The situation has changed now in Europe. In 1994 first steps were made towards developing an air traffic services Crew Resource Management programme and in 1995 the work started with a first task force of human factors experts, active controllers and training experts from across European States to look into the feasibility of what was then already called 'Team Resource Management' (TRM). The task force concluded that in fact TRM was feasible and beneficial and submitted Guidelines for developing and implementing TRM (EUROCONTROL, 1996).

The key objective of TRM is to develop the attitudes and behaviour towards enhanced teamwork skills and performance in ATM. Hence TRM aims to ensure the effective functioning of operational staff by helping them to use all available resources in time and as proficient as possible to reduce team work failures as a contributing factor in ATM related incidents and accidents.

TRM Prototype Course

The course developed with the support of active controllers provides a generic content and structure carefully selected and refined to be culturally acceptable balanced for the majority of nationalities and operational cultures. The modules are open for customisation and adaptation and including national examples i.,e. from incidents used in the training course and suit the learning needs of participants.

The prototype course consists of six modules (see Table 2 below) plus an introduction and a conclusion module. The course itself lasts for three days and is designed for 8 – 12 participants.

Table 2. *TRM Prototype Modules Content*

Team work (TW) – Typical characteristics of ATC related TW; negative impacts of behaviour on TW; character types in teams and impact on TW; team identity; safety issues related to TW; recognition and management of diversity in teams
Team Roles – Understand formal / informal hierarchies; attitudes towards authority (cultural impacts); strategies to avoid misunderstanding that leads to errors in the roles as a leader / follower; strategies to deal with submissive, aggressive and assertive behaviour
Communication (COM) – Functions of COM; understanding team COM related to safety; effective COM and effective intervention in ATM related situations; strategies to give / receive feedback and constructive criticism
Situational Awareness (SA) – Understand SA and the effects of high / low workload on SA; identify symptoms of team / individual loss of SA and strategies to prevent loss of SA; identify factors that have positive / negative effect on SA
Decision Making (DM) – contributing factors for effective DM; appreciate importance of situation and risk assessment skills for DM; appreciate concepts of shared problem models and resource management skills in team DM; structured DM
Stress Management – Identify job related stress factors; stress – what it is and how it affects work and team work; stress coping strategies; develop skills to recognise and cope with stress situation in teams

TRM trainers (facilitators) are trained in facilitation techniques which include self-presentation, mini-lessons, interactive lessons, introducing, summarising and conclusion techniques for discussions etc. The course material includes a facilitator and a participant handbook and video scenarios for some modules.

TRM Customisation and Implementation

ATM organisations that want to implement TRM in their operational and training environment can customise the modules using their own resources, the support from Eurocontrol experts or from external companies of their own choice. Guidance material is available to facilitate this customisation work (Woldring & Amat, 1998.

The cultural differences between States are often substantial and require a sensible approach. Local examples on incidents, the use of local language, humour, stories and staff are important and increase acceptance, awareness and impact in learning and actual application. This stresses the importance of TRM as learning rather than a teaching experience. TRM is a learning process that aims to positively impact on actual behaviour.

The implementation in ECAC States is still progressing. TRM users exchange their experience and expertise in a TRM User Group that consists mainly of air traffic controllers as TRM facilitators.

Recently tools complementary to TRM are available, one is the 'Behaviour Oriented Observation Method' (BOOM). Its aim is to train TRM facilitators and training instructors in objective, reliable and valid behaviour observation and feedback methods for non-technical skills in the TRM context. This is an additional step forward to increase the impact of TRM in practice.

In summary: TRM has proven to be widely accepted in European ATC now. It has helped to increase the awareness of human factors in general in ATM operations and increased the understanding of individual, group and cultural aspects in teamwork related behaviour. This also has helped to better understand why human errors can occur as a result of poor TRM.

TRM is now recognised as an important human factor in safety management. The Strategic Safety Action Plan (SSAP) established as a reaction to the Ueberlingen mid-air collision in 2002 recognises the importance of teamwork and team culture. The requirements set in the SSAP are mandatory for safety regulators and ANSPs in Eurocontrol member States. They need to ensure safety awareness, shall establish a safety culture, attitudes and behaviour amongst air traffic controllers through the implementation *inter alia* of measures in line with TRM. They are required to allocate the required resources for it and to report about implementation.

The long road for TRM to become practice would not have been achieved without the continuous and persistent efforts that human factors experts across Europe have invested in this area. This has fostered a better understanding of cultural diversity in teamwork in European ATM but also to bridge differences in local team and safety cultures. TRM has hence an important role to play in the future as a means to support change and transition and the merging of cultures in cross border Functional Airspace Blocks (FAB), the integration of teams in case of merged centres or units or centralises services.

The development of TRM is continuing on a communal and collaborative basis in the TRM User Group. Two new Modules have recently been developed. One is on 'Error Management' and is expanding on human error in teams and teamwork, the other is on 'Impacts of Automation' and addresses the cognitive impacts of current and future automation on individual and team performance, decision making and actions. More modules on the integration of teams and team cultures are planned.

CISM – Critical Incident Stress Management

CISM in short is a structured approach to assist people who have experienced an abnormal or traumatic critical event and react with strong personal emotions. The after-effects of critical incident stress can be substantial and long-lasting and can pose a danger to the well-being and performance of individuals and can even create a concern for safety.

The CISM work done in Eurocontrol started with a small booklet – called 'module' on CISM (EUROCONTROL, 1997a) which gave guidance for setting up of CISM in three phases:

- Information Phase: making aware and provide information on critical incidents, reactions how CISM support would come into force.
- Training Phase: Provide detailed information on CISM and the training of volunteers that would assist colleagues after critical incidents.
- Support Phase: Services and support that can be given to the persons concerned after the event.

CISM techniques involve a variety of methods and approaches and include:

- Early intervention – Don't wait after the incident happened!
- Use group dynamics – If more people are involve, get them together to speak and moderate the impact of the critical incident.
- Verbalisation, emotion ventilation etc.
- Debriefing and Defusing – Use this method to help to relieve emotions in a constructive and structured way.

Benefits and Lessons Learned in CISM Application

CISM case studies from various users demonstrate the benefits of investing in CISM. As in TRM customisation of the CISM approach to local and organisational culture and the working conditions and the use of peers as CISM experts are important aspects. The reports from existing schemes and from recent cost-benefit analysis on CISM indicate that CISM helps controllers to cope with the stress and return to work more rapide after critical events. It also reduces the risk of post-traumatic stress disorders that could lead to long-term sickness or even incapacity to continue to work in operations. The return on investment is reported to be positive.

As a high-priority action for safety related human resources in ATM ANSPs are required to implement CISM as an integrated element of their safety management system.

CISM's main strength is that it is a peer support model which has the effect of changing attitudes to critical incidents and the ways these are regarded. Organisations have noted more openness to the discussion of incidents and errors as a bi-product of CISM programs. As with HERA and TRM CISM has the potential for the future changing European environment in terms of making aware, address and manage critical incidents in a safety critical but fast changing ATM working environment with potentially high incident risks.

FEAST – A European Selection Test Package for Controllers

Background

Compared with pilot selection, the selection of candidates for controller training has fallen short in a number of major respects: Task and job analysis, selection development, test validation and use of best practice and standards. Common European or even international developments are still rare or non-existent, with only a few exceptions.

EUROCONTROL (1997b, 2000, 2001) gave detailed information on the situation in controller selection in the ECAC States based on detailed surveys done over the years continuously showing that around 50% of States could not provide appropriate results on the main psychometric properties (reliability, objectivity and validity) of tests they were using. The situation did change significantly due to improvements achieved in some States that recruit higher numbers of controllers annually. However, States that select and train only small numbers of controllers per year report that they found it difficult to comply with some of the Guidelines that Eurocontrol had issued. (EUROCONTROL, 2001, 2002). The conclusion was, that implementation, validation and maintenance of psychometric sound, complete and effective selection tests was neither practically nor financially feasible for some selection users. And that this had an impact also on training success, a low credibility of the tools used and has led to low stability and length of use of the tools leading to a lack of experience, validation possibility etc. and in fact into a vicious circle.

This situation was eventually addressed. The need for a common and advanced European development that would be based on tests that had demonstrated validity became an issue. From an initial reluctance against common standards, methods, guidelines and tools in 1995 emerged a situation of openness and support towards

- European communal efforts;
- Harmonisation of approaches;
- Establishment of enhanced quality and standards and benchmarks in test use and application;
- Common new test developments and even
- Common establishment, maintenance and validation of appropriate selection tools and methods for European wide use.

The 1999 Eurocontrol selection seminar strongly recommended to work together towards establishing a commonly-used selection system. 'Means and options should be investigated ... to acquire and / or develop a *European Controller Test Battery* that could be used in those ECAC States that are in need of this, especially the smaller States' (EUROCONTROL, 1999, p. 157).

In parallel to these European activities, participants at the 'International Air Traffic Controller Selection Conference' organised by FAA and held in Oklahoma City in the same year 1999 proposed to create an international working group of experts in ATCO selection. This group should openly exchange experience and data and share tools and ideas for mutual benefit and advancing developments in selection and especially in the cross-cultural validation of new test developments.

FEAST Development Objectives

FEAST was developed as a European joint venture with the objective to provide a basic, easy to administer and manage controller selection test option which reflects future impacts (e.g. as a result of changes in technology and the work environment) and enables customised implementation and use in Member States.

The test package should be flexible enough to be customised for use in different European countries and for the current and future tasks of the ATCO. Typical users for FEAST are States that recruit only small numbers of trainees annually and lack in-house expertise for own development and validation. Under these circumstances it will take long to obtain a sufficient validation sample size. Generally users would be interested if the cost-benefit ratio for an own fully fledged selection methodology would indicate prohibitive high efforts and costs.

FEAST Development Milestones

The list of milestone that had to be passed during the development cycle can be summarised as follows:

1. Demonstrate feasibility and viability of the FEAST concept.
2. Gain initial commitment from potential users across Europe.
3. Establish a controller job requirement model as the basis for development.
4. Gather, evaluate and select potential tests and methods for FEAST package in line with the model.
5. Compose a consistent test package.
6. Establish test delivery platform.
7. Establish comprehensive, consistent and licence and test user agreements and privacy policy for test takers.
8. Adapt and establish tests and scores for a consistent and easy administration and use.
9. Perform Quality Assurance and Standardisation of all procedures and tools.
10. Develop, investigate and validate a common criterion for initial and long-term validation of FEAST, the 'Behavioural Observation Scale' (BOS) for trainees and controllers.
11. Perform initial validation on multiple samples and groups of candidates / trainees / active controllers across Europe.
12. Establish FEAST as a service (service feasibility and implementation).

This paper can only highlight some aspects of the outcome of the FEAST programme in regard to concurrent validation findings and implementing FEAST as a service. (See for further details Rathje, & Golany (2003b) and Rathje, Golany & Eissfeldt (2004a and b))[6].

The FEAST Package

The FEAST package consists of tests composed into two assessment phases plus one optional assessment module:

Phase I Tests: Six web-based cognitive and knowledge tests running on PC linked to the Internet and include an English listening test. Standardised test results and a composite score are used for screening of candidates.

Phase II Test: A complex, dynamic multiple-task test administered on a standalone PC. The test simulates procedural control using flight strip data. Candidates are trained before taking the test using an integrated computer based training module. Candidates for Phase II testing are pre-selected, based on their results in the Phase I.
Optional Assessment Module: A Situational Interview (SI) paper-and-pencil format.

Initial Validation of FEAST in Cross-cultural Samples

Samples - The initial validation trials were conducted in 2002 – 2003 using a variety of samples and groups across nine European States. A total sample of 579 applicants, trainees and controllers were tested. The variation between samples in terms of age, gender, sample size and composition, selection stage etc was big and could not be influenced. The initial validation samples had hence to be used as given.

Criterion & Predictors - For Trainees and controllers concurrent criterion data was gathered using the common criterion method 'Behavioural Observation Scale' (BOS)[7]. The predictors were FEAST test scores and composites. They were correlated with the BOS and other training performance criteria in those samples where this criterion data was made available.

Restriction of range - It is well known that as participants in a concurrent validity study are a selected group of those persons who are actually and/or potentially able to do the job the variance in the predictors (and in the criterion) is likely to be smaller than that in a group of applicants for the job. The effect is a reduction of the size of the correlation coefficient. To estimate the correlation for the population, the standard deviation of predictor test scores in the applicant population, respectively in a sample from the population is to be known. For FEAST some data was available from real applicants in selection testing under standard conditions. Only these results are reported here in brief.

[6] The papers summarise the approach and methods adopted in FEAST validation in conjunction with the opportunities and challenges and problems encountered in cross-cultural validation, the development and validation of the concurrent Behavioural Observation Scale (BOS) method used to establish a common criterion and the detailed reliability and validity data both of the predictors and the criterion development and validation (the FEAST Behavioural Observation Scales, BOS) and on the controller job requirement model.

[7] Concurrent validation as a means to establish the validity of test scores if a full and comprehensive long-term, predictive validation approach cannot be performed, or as a first step in a more comprehensive validation process, as was the case with FEAST. As such, concurrent validation is not an alternative to a predictive validation but can offer an independent measure of test validity.

Phase I Test Results - The coefficients reported in Rathje, & Golany (2003b) were computed on (restricted) samples of Trainees and Controllers. A comparison of the (standardised) test scores and composite scores between candidates and trainees showed that the variance of the scores of trainees for example were between 55% and 90% of the variance in the candidate group. Corrected (adjusted) correlations were computed based on the standard correction formula (Hunter & Schmidt, 1990).

The correlation between the Total Test Composite and FEAST Criterion (BOS Summary Score of a total of 35 Items) was r =.296 (p< 0.05) in one sample of n=55 Trainees. The corresponding adjusted correlation based on the correction formula for range restriction is r = .42 (p< 0.01) for the same sample (n=55)[8].

The Controller samples were quite different one from another - more than the trainee ones – as they varied more in their age range, level of motivation for taking the FEAST and the level of consideration given by their supervisors for completing the standard BOS criterion scales. When using one sample of n = 24 ATCOs from one location with a known selection ratio and selection methodology, age range and homogenous, reliable and complete criterion data, the restriction in the range would have been even higher. Although the correlation between the FEAST Composite Score and BOS Summary Score is significant (r = .46, p < .05) (adjusted r = .57, p < .01) the sample size (n=24) is too small for drawing conclusions.

The correlation between FEAST predictors (composite score) and other training criteria, for example, 'Course Overall Final Pass Mark' (training score at the end of Initial Training) from one trainee sample (n = 46) - where the selection ratio is 12.5% - to be **r = .36** (p< 0.05).

Phase 2 Test Results - Regarding the complex work sample test, the correction formula for range restriction[9] was applied on a trainee sample for 3 scales of the BOS: "BOS Summary Score" (35 Items, as above), "Teamwork" and "Working under Stress". Here, for example, the corrected correlations for a predictor of this test called 'number of correctly identified opposite conflicts' with the BOS total score, was r = .41 (p< 0.01) for a trainee sample (n=43). The same predictor's adjusted for restriction correlation with the BOS score of 'Working under Stress' was r = .39 (p< 0.01) (n=50 trainees).

The correlations between one composite score representing the performance in regard to updating / inserting and ordering flight strips (Total Performance in one performance criterion – Strip Order) with the three BOS scales are given in Table 3 (adjusted r in brackets):

Table 3. *Adjusted and unadjusted correlation between criterion and predictor scores in FEAST Phase Testing - trainee samples*

BOS – Summary Score	**. 44** (.54**)** (n = 43 Trainees)
BOS – Teamwork	**.43** (.54**)** (n = 50 Trainees)
BOS – Working under Stress	**.37** (.47**)** (n = 50 Trainees)

All correlations are significant at the < .01 level.

FEAST Cross-cultural Validation and Testing Conclusions

The results of the various studies clearly demonstrated the challenges in a cross-cultural, common approach in validation of controller selection tests and the impact of sample size, age and composition of the validation samples, the restriction of range due to failures in training and other aspects that have or can have detrimental effects on the results.

Use of a Common Criterion Measure (BOS) - As regards the BOS criterion measure, the studies demonstrated the reliability, validity of the BOS as well as the relevance and need of a common criterion measure. The findings in some samples however also give warnings as regards the need for appropriate training material and calibration training in using the BOS (or other measure) of assessors i.e. training instructors or supervisors. The differences in training, trainee assessment methodology and culture are big. Important items for consideration and countermeasures for the future, long-term validation of FEAST are the use of behavioural anchors for all scales of BOS and reasonable training of assessors.

FEAST Predictor Tests - The tests and the various composite indices developed for use in selection decision-making, the findings from the initial validation study showed that the scores are sufficiently reliable and stable across samples. The results also demonstrated that despite the low

[8] This result was cross-validated using a different composite score which allowed the analysis of a sample of n = 81 Trainees. Here the correlation was r = .27 (p < .05) (not adjusted) and r = .45 (p < .01) adjusted.

[9] The variance in the trainee sample compared to a sample of (pre-selected) candidates that took this test as the Phase II test was less restricted as expected and was between 70% - 120% of the candidate sample

variance in the FEAST BOS criterion (especially in the Controller samples, significant and stable correlations were found between test scores and the criterion in samples of trainees and qualified controllers.

In Summary: The study efforts already now demonstrate the progress in efficiency and benefits that can be reached by the application of a proper, valid selection procedure and by a combined validation effort across various European air traffic controller training schemes and the establishment of common 'European' norms. It is made clear that this can only be achieved if a common, collaborative and harmonised approach is adopted and quality criteria and standards are shared and actually met. Whether this is feasible to achieve is still an important challenge in the establishment of a Pan-European FEAST service.

FEAST Service Feasibility and Implementation

FEAST since the beginning of 2004 has progressed into a 'FEAST Service Planning and Feasibility Phase'. The aim of this phase is to establish the viability of the nature and scale of such a service offered by Eurocontrol to ANSPs that wish to use FEAST during a pilot Service. During 2004 - 2005, the viability of the service delivery is established to prepare a full business case that will assist decision-making regarding the introduction of a full FEAST service. During this period, FEAST is tested under 'live conditions' on real applicants and test data is gathered as an input into a longer-term predictive validation study and to establishing common norms.

The current experience in now seven different States where FEAST has been implemented is very promising. The implementation requirements for FEAST include the training of administrators in standard test administration, FEAST application and installation and technical requirements. FEAST recruiters are specially trained in the valid use and interpretation of test scores. Standards in regard to the test environment are observed during local installation visits. Local customisation and the integration of FEAST into current existing selection and recruitment methods are essential.

FEAST - Potential of Improving Selection in Europe

FEAST implementation and validation findings so far demonstrate the feasibility of a service across different cultures. FEAST

- Offers a valid and scientifically sound test battery;
- Meets agreed Eurocontrol guidelines in selection and recruitment;
- Enables quicker validation, a bigger sample and proper predictive validation in the lon-term;
- Reduces development, validation and maintenance and upkeep costs;
- Ensures high quality and standards in testing and selcting candidates for ATC training;
- Fosters efficiency and effectiveness in selection;
- Includes a built-in continuous improvement.

FEAST is a project geared to continuous, ongoing improvements and maintenance. Continuous improvement is crucial for the long-term viability and sustainability of FEAST together with users and scientific partners in development.
Further developments involve parallel test versions and new tests. One recent example is the development of methods for a valid and reliable testing of English language proficiency in all performance areas of the new ICAO requirements.

References

Barbarino, M., Isaac, A., Krois, P. Rodgers, M. & Van Damme, D (2004). New technique improves analysis of human factors in ATM incidents. *ICAO Journal, 58, (8), 16-18, 24.*

EUROCONTROL (1996). *Guidelines for developing and implementing TRM* (Brussels: Eurocontrol.

EUROCONTROL (1997a). *Human factors module – Critical incident stress management*. Brussels: Eurocontrol.

EUROCONTROL (1997b). *Information on Available and Emerging Selection Tests and Methods for Ab-initio Trainee Controllers.* Brussels: EUROCONTROL.

EUROCONTROL (1999). *Proceedings of the First EUROCONTROL Selection Seminar - Current and Required Future Selection Work and Methods in the ECAC Area*. Brussels: EUROCONTROL.

EUROCONTROL (2000). *Update of Information on Selection Tools and Methods for ab-initio Trainee Controllers*. Brussels: EUROCONTROL.

EUROCONTROL (2001). *Guidelines for Selection Procedures and Tests for Ab-initio Trainee Controllers*. Brussels: EUROCONTROL.

EUROCONTROL (2002). *Selection Tests, Interviews and Assessment Centres for Ab-initio Trainee Controllers: Guidelines for Implementation.* Brussels: EUROCONTROL.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Beverly Hills, CA: Sage.

Rathje, H. & Golany, Z. (2003a). FEAST – Milestones in the Development of a Pan-European Project. In: *FEAST and More – Progress in European Controller Selection Development. Proceedings of the 2nd EUROCONTROL*

Selection Seminar 16-18 November 2003 (in press).

Rathje, H. & Golany, Z. (2003b). FEAST Initial Validation: Summary. In: *FEAST and More – Progress in European Controller Selection Development. Proceedings of the 2nd EUROCONTROL Selection Seminar 16-18 November 2003* (in press)

Rathje, H., Golany, Z. & Eissfeldt, H. (2004a). Pan-European Selection Test Battery for Air Traffic Control Applicants, In: K.-M. Goeters (Ed.), *Aviation Psychology: Practice and Research*, Chapter 10, pp. 171-201, Aldershot, Hampshire: Ashgate.

Rathje, H. & Golany, Z. (2004b). *The Challenges and Benefits of a Combined Validation Effort across States*. Paper presented at the 23rd EAAP Conference, Sesimbra:Portugal.

Wiener, E.L., Kanki, B.G. & Helmreich, R.L. (1993). *Cockpit resource management*. San Diego: Academic Press.

Woldring, M. & Amat, A.-L. (1998). TRM in European ATC: *Customisation of a prototype course*. Paper for the 23rd EAAP Conference, Vienn, Austria.

Woldring, M., Van Damme, D. Patterson, I & Henriques, P. (2005). Team Resource Management in European air traffic control: results of a seven-year development and implementation programme. In: B. Kirwan, M. Rodgers, and D. Schaefer. (Eds) *Human Factors Impacts in ATM*. Ashgate, Aldershot (in Press).

PILOTS, AIRSPACE COMPLEXITY, AND STRATEGIC CONFLICT AVOIDANCE

Victor Riley, User Interaction Research and Design, Inc., Point Roberts, Washington
Edmundo Sierra, Titan Corporation, Mays Landing, New Jersey
Gano Chatterji, University of California Santa Cruz, Moffett Field, California
Richard Mogford, **Walter Johnson**, and **Parimal Kopardekar**, NASA Ames Research Center, Moffett Field, California
George Lawton, Raytheon STX, Moffett Field, California

Some future air traffic management concepts seek to place more separation responsibility on the pilot in order to achieve greater aircraft operating autonomy. Separating one's own aircraft from others in something other than a see-and-avoid environment, however, would pose fundamentally new demands and challenges for pilots, and it is likely that new automation and display tools would be needed. Ideally, an automated strategic conflict avoidance system would behave consistently with pilot expectations and take pilot interests into account when suggesting resolution strategies. It might also recognize situations that pilots may have difficulty detecting and resolving on their own. At this time, little is known about how pilots perceive airspace complexity in self-separation tasks. In this study, we used a Cockpit Display of Traffic Information (CDTI) with an embedded strategic conflict avoidance aid to help fourteen commercial transport pilots detect and resolve a series of strategic conflict situations. We then assessed their performance with and without the aid, recorded and analyzed pilot ratings of aid effectiveness and usability, and used a neural network model to associate complexity ratings with airspace characteristics to determine which sets of characteristics most heavily influenced pilot perceptions of airspace complexity. The results of this analysis provide insight into what aspects of airspace configuration may have the greatest influence on pilot perceived workload and difficulty understanding conflict situations.

Introduction

Several emerging concepts for future air traffic management systems seek to transfer some, or all, of the responsibility for aircraft separation from air traffic controllers to pilots. In most concepts, such as the "Free Flight" concept (RTCA, 1995), this is done to grant airlines and pilots more autonomy, under the assumption that this will lead to more efficient routing and allow operators to optimize their routes or the flow of traffic within their fleets.

For airline and instrument pilots, this will represent a new set of responsibilities and is likely to add to their workload. Furthermore, flight management responsibilities may sometimes make it difficult or impossible to also attend to self-separation responsibilities. This suggests that some form of automated assistance will likely be needed.

In this study, we investigated the usefulness and usability of a Cockpit Display of Traffic Information (CDTI) with an embedded automatic strategic conflict detection capability, coupled with a route planning aide that assessed the presence of conflicts associated with modified routes, in a variety of strategic conflict situations. Recognizing the potential for mismatches of interests, solutions, and expectations between the pilot and automation, we also used a neural network model to better understand what aspects of the airspace, based on the positions and trajectories of nearby traffic, most contributed to pilot perceptions of airspace complexity.

In particular, the questions we were interested in included:

- What characteristics of the airspace (positions and velocities of other aircraft) affect pilot perceptions of airspace complexity?
- How much benefit would a decision aid be in detecting and resolving conflicts?
- How readily would pilots accept and use such an aid?
- What characteristics of the airspace affect the ability of pilots to reliably detect and resolve conflict situations without help?
- How much complexity can pilots reliably handle before decision making and route replanning performance start to deteriorate?

The results of this study should help guide the development of automated route planning and conflict resolution aids and ensure that such aids adequately account for pilot interests and expectations. It should also help guide the development of airspace management procedures involving aircraft with self-separation capabilities.

Method

This study brought two bodies of prior work together to support the effort: a CDTI developed at NASA, and prior work using neural networks to understand how air traffic controllers are influenced by airspace complexity factors.

CDTI

We used a CDTI/route planning aid (hereafter referred to simply as the "aid") that was already under development at NASA (Johnson, Battiste, Delzell, Holland, Belcher, & Jordan, 1997). The display, shown in Figure 1, depicts own aircraft position at the lower center of the display and nearby traffic represented as chevron symbols. These symbols are green when the other aircraft are below own-ship, white when at the same altitude, and blue when above. When the system detects a conflict, ownship and the conflicting aircraft turn amber, an amber connecting line is drawn to show the projected conflict position, and an audible alert is given. In addition, aircraft that may come close to own-ship but do not currently conflict are shown in amber outline without an alert to help the crew monitor traffic that might merit special attention.

Figure 1. *CDTI showing a projected conflict*

In addition to displaying relative traffic positions and conflict status, the aid can also display aircraft information and flight plan intentions. The user can also select a pulse display feature in which the positions of aircraft along their vectors are projected into the future. This allows the user to compare the relative positions of aircraft into the future, and determine what the order of aircraft arrival at a crossing point would be.

The CDTI also includes a real-time route planning module that allows the pilot to adjust the path while receiving real-time feedback about whether the adjusted path would be free of conflicts. The pilot can "grab" the path with a cursor and move it in either direction, or place a waypoint on the path and assign an altitude change to the waypoint. This capability made the CDTI a good platform for performing the experiment manipulations required for this study.

Airspace Complexity Factors

A number of prior studies (Chatterji & Sridhar, 2001; Kopardekar & Magyarits, 2002; Kopardekar, 1997; Mogford, Guttman, Morrow, & Kopardekar, 1995) have examined the effect of airspace complexity factors on air traffic controller perceptions of airspace complexity. Various sets of factors have been introduced by a variety of authors, but they usually include parameters associated with the number of aircraft in an area, the number of aircraft within an altitude band, the number of aircraft changing trajectory either laterally or vertically, the presence or absence of conflict conditions, the angle of convergence in a conflict, and others. The number of these measures suggested by various authors jumped after the RTCA free-flight concept (RTCA, 1995) because this concept included a notion of "dynamic density" characterized by airspace complexity factors; a given airspace would be under either free-maneuvering rules or positive control depending on its dynamic density.

We surveyed the collection of airspace complexity factor lists that had been compiled, eliminated those factors that could only relate to ground-based control (and therefore were not relevant to self-separation), and then eliminated factors that were essentially identical to arrive at a list of potentially relevant and unique factors. We then collected these factors into 21 sets for use with the neural network.

Experiment

Fourteen commercial pilots, all male and all with glass cockpit experience, participated in the experiment, which was performed on a laptop computer. The pilots were asked to resolve fourteen conflict situations, which had been designed to cover a range of difficulty levels from very low to very high, and to include a variety of conflict types (two vs. multiple aircraft involved, head-on conflicts vs. shallow-angle conflicts, and conflicts with aircraft that were changing altitude to or through own altitude). Prior to performing in the experiment trials, the pilots were given a short training presentation about the nature and procedure of the study, filled out demographic questionnaires, and completed six training trials.

Each trial began with the traffic configuration appearing, detected conflicts shown, and the display freezing so the pilot could study the situation. A rating box was displayed so the pilot could rate the complexity of the situation on a three point scale. The pilot could examine the flight and flight plan information for any aircraft on the display, and did not need to enter the complexity rating until fully understanding the situation. We measured the time from scenario start until the pilot entered the complexity rating in hopes that this time measure could serve as an objective measure of complexity (under the theory that it would take more time to understand a more complex situation). As it turned out, there was no significant correlation between the time required to enter the complexity ratings and the ratings themselves.

After the pilot entered the complexity rating for a given scenario, the display would resume motion and the pilot would have the opportunity to adjust the route to resolve conflicts. In half of the experiment trials (counterbalanced for scenario and order), the pilot would be provided with real-time feedback from the aid about whether the adjusted route had resolved the initial conflicts and whether it had created any new ones. In the other half, this feedback was not given; the initial conflict continued to be depicted and the pilots were asked to judge on their own whether the adjusted route was conflict-free.

When the pilot was satisfied that the adjusted route was free of conflicts, he would enter the route into the system, which would then provide feedback that the new route had been activated. Then, another rating box was displayed, this one asking the pilot to rate the difficulty of resolving the situation on a five point scale. Once this rating was provided, the experiment would move on to the next trial (or end).

After completing the experiment trials, the pilots filled out a survey covering their attitudes regarding the usefulness and usability of the aid. The pilots were paid $100 for their participation.

Analysis

The list of measures included:

- the complexity ratings
- the time required to enter these ratings
- airspace configuration at the time of these ratings
- the difficulty ratings
- total completion time for each trial
- whether any conflicts remained at the end of each trial
- pilot ratings of aid usefulness and usability.

We tested the effects of the aid on complexity and difficulty ratings using repeated-measures regression analyses as well as tests of differences in regression coefficients. Repeated-measures Analysis of Variance (ANOVA) methods were used to determine the effects of the aid on the accuracy and total time of resolving conflicts. Finally, we evaluated the pilot ratings (on a seven point scale) of aid acceptability by inspection.

The neural network analysis was more involved. As mentioned earlier, the airspace characteristics, as represented in terms of the selected complexity measures, were recorded at the time of the entered complexity ratings. This allowed us to associate the complexity measures with the subjective complexity ratings. We then trained a neural network to reproduce the aggregate complexity ratings through an iterative feedback process, as shown in Figure 2.

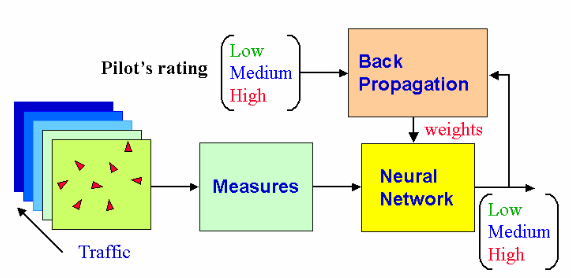


Figure 2: *The neural network used iterative back propagation to "learn" how to produce complexity ratings representative of the pilot ratings*

Through this iterative back propagation process, in which the network attempted to reproduce the aggregate pilot complexity ratings with each of the 21 sets of complexity factors, we were able to determine which sets of factors produced the best match between pilot ratings and the neural network outputs. In other words, we determined which set of factors caused the network's behavior to most closely match that of the pilots.

Results

In summary, the results demonstrated that the aid improved pilot avoidance of conflicts and was generally accepted and liked by the pilots. We were also able to identify a set of eleven airspace complexity components out of the original 21 that appear to most heavily influence pilot complexity ratings.

One of the most informative pilot performance results had to do with the comparison of the complexity and difficulty ratings for each scenario. This is because of the possibility that the aid might be able to make complex situations relatively easy to resolve. To facilitate this comparison, we used a three point scale for complexity ratings and a five point scale for difficulty ratings to discourage subjects from merely repeating their complexity ratings, given at the start of the scenario, in the difficulty ratings given at the end.

The reason for suspecting that the aid might make complex situations easy to resolve had to do with how the aid transformed the nature of the task. Without the aid, the pilot had to mentally visualize how the airspace situation would change over time. With the aid, the pilot had to merely adjust the route until none of the aircraft symbols and vectors were yellow. This transformed the task from a complex, multidimensional visualization involving multiple targets to a simple binary judgment.

Indeed, we found that the correlation between these two ratings dropped significantly when using the aid, from $r = 0.76$ with the aid off (in the trials where no assistance was provided during route adjustment) to 0.62 with the aid on (two-tailed alpha = .05). This demonstrated that the aid effectively decoupled the difficulty of resolving the scenario from the conceptual complexity of the scenario.

We also found that subjects who resolved conflicts first using the aid rated the overall complexity of all scenarios as more complex than those who resolved conflicts first without the aid ($F(1, 12) = 5.00$, $p < .05$). This suggested that using the aid informed the subjects about the true complexity of scenarios, perhaps by showing them conflicts that they would not have otherwise noticed. This may have caused them to better appreciate the complexity of scenarios they later attempted to solve without the aid.

As expected, subjects resolved conflicts more accurately when using the aid (88% resolved) than they did without the aid (77% resolved). (With one statistical outlier removed, these figures were 90% and 76% respectively.) However, it took pilots longer to resolve conflicts when using the aid. This may reflect the absence of feedback when attempting to resolve conflicts without the aid; without information that the adjusted route was conflict-free, subjects may have entered the new route more quickly than they would have with feedback that there were still conflicts present.

We used a seven point scale to measure pilot opinions about the aid's usefulness and usability. In general, pilots gave the aid favorable ratings for both. They indicated that they would like to use the aid in a free-flight environment, but they expressed concern about the proposed changing roles of air traffic controllers and pilots; several of the subjects commented that they would prefer to retain positive ground control, but that if they had to operate in a free-flight environment, they would value the assistance of the proposed aid. They also indicated that they were not confident resolving conflicts without the aid (nine subjects expressed lower than neutral confidence, three higher than neutral, and two neutral).

In order to learn how to approximate the pilots' complexity ratings, the neural network had to be given an aggregate set of ratings (low, medium, high) that represented the "consensus" rating of the group. To do this, we calculated, for each scenario, a weighted average rating with a floor function to match the weighted average to the rating scale. Taking this weighted average as the aggregate rating for each scenario, we were able to assess the representativeness of the aggregate ratings by calculating the proportion of pilots whose ratings agreed with the aggregate, for the three levels of rating. This is shown in Table 1.

	Pilot Ratings		
Aggregate	Low	Medium	High
Low	69.6%	28.6%	1.8%
Medium	30.5%	62.8%	6.7%
High	7.2%	35.7%	57.1%

Table 1. *The proportion of pilots whose ratings agreed with the aggregate ratings*

We then used this aggregate rating set as the criterion to be approximated by the neural network through the back propagation process. In general, the network solution stabilized after about a thousand iterations. The proportion of neural network ratings that matched the pilot ratings is shown in Table 2.

	Neural Network		
Pilot	Low	Medium	High
Low	68.3%	30.8%	0.9%
Medium	14.4%	81.7%	3.9%
High	0%	43.5%	56.5%

Table 2. *The proportion of neural network ratings that agreed with the aggregate pilot ratings*

Table 2 shows that the neural network did a very good job of emulating the pilot ratings. With the neural network trained to behave approximately as the pilots did, we examined how well it performed with the various sets of airspace complexity components, reasoning that the set of components that gave the best match between the neural network and pilot ratings would best represent the set of influences on the pilot's own perceptions.

Several sets of components scored relatively well, differing in how well they matched either the high or low ends of the scale (that is, some sets closely matched the low complexity ratings but did less well on the high, while others did the reverse). One set that showed the best balance across the scale and contained a relatively sparse number of components included the following components:

- the total number of aircraft in the scenario
- the number of climbing, cruising, and descending aircraft
- measures of horizontal and vertical proximity
- amount of time remaining before conflict
- the ratio of the standard deviation of speed to the average speed of aircraft in the scenario
- the number of unique alerts ongoing
- the presence or absence of an alerting state
- the presence of shallow angle conflicts (which are particularly difficult for pilots to recognize and project).

For this set of components, the neural network matched the pilots' aggregate "low" rating 68.3% of the time, the "medium" rating 81% of the time, and the "high" rating 52.2% of the time.

Conclusions

These results provide an initial step toward understanding how pilots conceptualize the local airspace in strategic conflict situations, and may help us better understand what capabilities and behaviors they will expect and need in a strategic conflict avoidance aid for a free-maneuvering environment. Ideally, these and the results of following studies will help designers compensate for known pilot performance weaknesses in such situations (such as poor ability to recognize shallow angle conflicts and to visualize conflicts involving aircraft with changing altitudes). They should also ensure that future self-separation aids take pilot interests and expectations into account, thus avoiding potentially surprising behaviors in potentially dangerous situations.

In the next steps for this work, we hope to add real-world maneuvering constraints such as weather and restricted airspace, and introduce traffic with changing flight plans, dynamic maneuvering, and possibly unreliable intent information. We would also like to compare pilot solutions to such conflicts with optimum solutions to better understand pilot strengths and weaknesses in such circumstances, determine at what levels of complexity pilot performance breaks down (and automation is required), and how pilots can effectively manage failures of the conflict aid in complex traffic environments. We hope that these results will inform not only future technology development in this area, but also the development of airspace management and flight deck procedures in free-maneuvering environments.

References

Chatterji, G. B., & Sridhar, B. (2001). Measures for air traffic controller workload prediction. (AIAA-2001-5242). Proceedings of First AIAA Aircraft Technology, Integration, and Operations Forum. Los Angeles, CA: Aviation Week Aerospace Expo.

Johnson, W.W., Battiste, V., Delzell, S., Holland, S., Belcher, S., & Jordan, K. (1997). Development and demonstration of a prototype free flight cockpit display of traffic information. Proceedings of the 1997 SAE/AIAA World Aviation Congress. Anaheim, CA.

Kopardekar, P. (1997). Dynamic density metric variables. (Federal Aviation Administration, William J. Hughes Technical Center, National Simulation Capability Program.)

Kopardekar, P., & Magyarits, S. (2002). Dynamic density: Measuring and predicting sector complexity. Proceedings of the 21st Digital Avionics Systems Conference, IEEE.

Mogford, R. H, Guttman, J. A., Morrow, S. L., & Kopardekar, P. (1995). The complexity construct in Air Traffic Control: A review and synthesis of the literature. (DOT/FAA/CT-TN-95/22). (Atlantic City: FAA Technical Center.)

RTCA. (1995) Final report of RTCA Task Force 3: Free flight implementation. Washington, D.C.: RTCA.

CHECKLIST USAGE AS AN INDEPENDENT VARIABLE IN STUDENT PILOT TASK PERFORMANCE ASSESSMENT

Dr. Vladimir N. Risukhin
College of Aviation, Western Michigan University
Kalamazoo, Michigan

Aviation safety statistics show that checklist utilization by pilots is one of many safety-critical aspects of flight operations. Flight training practice and experience in teaching student pilots to the principles of the multi-crew flight environment require addressing the checklist issue. Conventionally there have been two contradictory flight safety aspects relevant to checklist usage in flight operations. Flight safety standards require that checklists must be performed in-full during normal and non-normal flight situations. Conversely, checklists can be sources of pilot distraction from controlling the airplane that may compromise flight safety. A FAA approved Flight Training Device (FTD) was used to prove the possibility to measure student pilot performance during various checklist applications. This study is directed to finding specific correlations between different methods of checklist usage and the level of student pilot performance. The proposed methodology may be applied for research and improvement of various pilot training programs.

The Problem of Checklist-Induced Crew Errors

Checklists have taken a recognized position in complex human-machine systems operation. In aviation checklists secure execution of actions critically important from safety point of view. The "stimuli-reaction" activity is an extremely simplified explanation of the checklist utilization by a human operator. In aviation it may be correct only within specific phases of flight operational process such as, for example, preflight aircraft check. In many other flight situations checklists usage coincides with other important tasks performed by the flight crew because crewmembers must be included into the aircraft control process during the whole period of flight.

<u>Continuous control of the flight path is the primary task of the flight crew</u> because the aircraft can be a safe flying machine only within a rigid range of its flight path parameter values. All other flight crew tasks are subordinate to this vitally important activity. Other tasks performed by crewmembers simultaneously with the flight path control activity may distract them and induce crew errors. The checklist utilization may be such a distracting task.

A History of Conflicting Tasks Problem Solving

The problem of a human involvement into several tasks performed simultaneously attracted scientists since before this problem became recognized in aviation science and industry. A prominent psychologist William James (1842-1910) described the human attention distribution between different tasks and inevitable distraction from one task in favor of another: "Every one knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state which in French is called *distraction*, and *Zerstreutheit* in German" (James, 1890). Many decades later O'Hare and Roscoe (1990) have stated that two tasks while being simultaneously performed by a pilot may create different levels of a conflict between them. There may be a smaller conflict between tasks that are not competing for the same resources of pilot's attention (such as monitoring spatial orientation of the aircraft and listening to an air traffic controller) in comparison with tasks that require the same attention resource.

The history of aviation development has confirmed usefulness of checklists for ensuring safe flight operations. Nevertheless, there are evidences proving that flight crew errors may be caused by the very process of checklist usage. Degani and Wiener (1990) cited a study showing that out of 169 airline crew distraction reports 22 were caused by checklist procedures. This problem represents a significant threaten to flight safety because pilots distracted from immediate controlling of the aircraft may inadvertently cause unacceptable deviations of the aircraft flight path parameters (altitude, heading, vertical speed, air speed etc) from their required values. Degani and Wiener (1990) also suggested that the checklist designer can decrease the probability of a checklist process interfering with other tasks by reducing the length of a given checklist.

Establishing of quantitative relations between human operators' psycho-physiological state and their performance is a prerequisite for successful obtaining of meaningful results in pilot performance research. A study of aircraft crew performance in flight path controlling process has confirmed correlations between flight operational factors influencing the crewmembers' psycho-physiological condition and deviations of actual flight path parameters from their required values (Risukhin, 1988).

Aviation regulatory agencies paid serious attention to the checklist utilization by crewmembers. The National Transportation Safety Board (NTSB) recommendations suggested that "more emphasis should to be placed on the use of checklists" (U.S. DOT FAA, 1995). The Federal Aviation Administration established requirements for crewmember activity and interactions during checklist utilization, such as "challenge-do-verify" and "do-verify" actions. These requirements are applicable to "normal" flight operations without any aviation equipment failures, as well as to "emergency" and "non-normal/abnormal" procedures when crewmembers have to cope with various equipment failures and operational abnormalities (U.S. DOT FAA, 2000).

Checklist Utilization Problems in Pilot Training

Professional pilots do not represent the only part of the pilot population that may reduce their level of performance due to checklist-induced distractions. *Checklist utilization by student pilots* during a flight training process may complicate development of trainees' flight control skills due to distractions generated by checklists. This problem may be significantly aggravated in multi-crew flight training environment when the aircraft controlling and checklist utilization processes performed simultaneously impose additional requirements on student pilots' attention, as well as communication and coordination of their actions.

Instrument flight proficiency and teamwork skills are the most critical characteristics of professional pilots required in contemporary aviation. Aviation education institutions have to develop student pilot skills in aircraft flight path instrument control with simultaneous utilization of checklists as integral parts of standard operational procedures (SOP). Properly organized line oriented flight crew simulation in flight training devices (FTD) which is an application of the line oriented flight training (LOFT) to aviation academic environment may help student pilots in simultaneous developing of their instrument flying and crew resource management (CRM) skills.

Conditions of Flight Checklists Utilization

The checklist utilization during every phase of flight is an operational requirement. It assures that the aircraft has been correctly configured for every phase of flight, and that all vitally important system control actions have been performed when needed. Checklist caused pilot distractions threaten the flight safety because they influence the crewmember's key psychological parameter - attention. Revelation, analysis, and neutralization of negative factors caused by checklist utilization in the process of pilot training and aircraft flight operations require a brief review of operational conditions in which checklists are used in aviation. Following factors characterize these conditions:

- the contemporary level of aviation technology that defines general design of aircraft, their cockpits and controls;
- the way of application controlling forces to aircraft controls (manual or automated);
- the method of visual information perception used by pilots (visual or instrument flight conditions);
- the type of air navigation system used for crewmembers actual flight path parameters perception;
- the phase of flight.

The Contemporary Level of Aviation Technology

Aircraft configuration requirements and their systems complexity are two characteristics of contemporary aviation technology that make checklists a compulsory instrument of the flight crew.

Different configurations of aircraft are used during specific phases of every flight (taxiing, take-off, climb, cruise, descent, approach, and landing).

Complexity of aircraft systems controlled by the flight crew requires optimally structured crew actions for systems activation, operations and control.

The Way of Aircraft Flight Path Control

The aircraft flight path control process is based on continuous comparison of actual and desired flight path parameter values. As soon as the aircraft controlling function (performed by the flight crew or the flight control computer) perceives a difference between the two values of a controlled flight path parameter, it develops and applies a control input to the flight controls to change the value of the controlled parameter. Two different ways of modern aircraft flight path control (manual and automated) define different ways of crewmembers' flight-path-controlling activity.

In the process of automatic flight path control pilots are responsible for preparation, activation, monitoring, and control of aircraft automation. They do not apply immediate inputs to the aircraft controls (elevator, rudder, ailerons and engine thrust control) but set desired flight path parameter values on the aircraft automatic control devices.

During manual control of the aircraft flight path one flight crewmember (the pilot-flying - PF) perceives the flight path parameter relevant information, processes it, and develops and applies control inputs to the aircraft controls. The second flight crewmember (pilot-not-flying - PNF) monitors the flight path parameters and helps the PF in maintaining of the required flight path by informing the PF about flight path parameter deviations.

The Method of Crew Visual Information Perception

Significant differences exist between flight crew information processing activities during visual and instrument flight conditions. Available sources of visual information define the crewmembers' attention distribution as well as their mental and muscular activity adequate for satisfactory control of the aircraft flight path.

In visual flight conditions pilots obtain a significant part of the information about the aircraft flight path by observing the aircraft attitude in the airspace through their visual perception of objects located outside the cockpit (the Earth horizon, terrain, aerodrome facilities, natural and artificial obstacles etc.). In the process of visual control of the flight path the crewmembers pay a relatively smaller part of their attention to incremental perception of the flight path parameter indications shown by cockpit instruments. In visual flight pilots use instrumentally perceived flight path parameters data (speed, altitude, bank, and heading) as important supplementary reference information needed for developing optimal control inputs.

During a flight in instrument flight conditions pilots obtain most of the flight path relevant information from cockpit instruments. In this case the instruments do not represent to pilots a visual picture, similar to that they obtain from outside the cockpit during a visual flight. Pilots must integrate fragmentary information perceived from various instruments to create a mental image of the actual flight. Then pilots compare characteristics of this image with required flight path parameters, and in case of discrepancies they develop and apply control inputs resulting from this comparison. Because of this fact the instrument flight control conditions require from the flight crewmembers a higher level of their cognitive activity than that required for a visually controlled flight.

The Type of Utilized Air Navigation System

Various types of air navigation systems used by crews for actual flight path parameter values perception provide flight path relevant data indications within various amount and precision ranges. A precision air navigation system like the Instrument Landing System (ILS) indicates exact deviations of the airplane from the intended flight path. Flight crew cognitive workload in the process of the precision system utilization is lower than the workload imposed by non-precision systems, such as Very high frequency Omnidirectional Radio range (VOR) or Non-Directional Beacon (NDB). Non-precision air navigation systems require from the flight crew a significantly higher level of cognitive efforts to calculate the flight path parameter deviations and to develop compensating control inputs.

The Phase of Flight

In addition to differences between requirements to crewmembers' attention amount and their actions distribution in manual, automatic, visual, and instrument flight, different phases of flight impose upon the flight crew different levels of workload. A higher level of human operator workload is usually conductive to human errors. Take-off, final approach, and go-around maneuvers are widely recognized as the most safety-critical phases of flight.

Rationale for the Research

In addition to the checklist structure optimization, proposed by Degani and Wiener (1990), a research may be suggested to pursue solving the problem of checklist interference with other important cockpit tasks in multi-crew flight operations.

A comprehensive study of the instrument flight control training process including checklist utilization within multi-crew operational environment is needed for better understanding and further improvement of student pilot performance.

Factors of development of the flight crew cognitive and muscular activity aimed at simultaneous flight path control and checklist utilization have to be identified, assessed, analyzed, and integrated into a unified model of flight crew information processing and control inputs development and application. The model has to be usable for optimal combination of the flight crew aircraft control and checklist utilization actions during flight crew training process as well as within aviation industry flight operational environment.

The achievement of the study goals requires a research aimed at finding of statistically confirmed correlations between the process of checklist utilization by the flight crew (an independent variable), and assessments of the crewmembers' flight path control performance together with their CRM skills (dependent variables).

The comparison of different methods of checklist utilization by student pilots used in the research may allow the choice of the most optimal sequence of instruction flow in flight crew training programs. *Optimization of student pilots training* in their simultaneous involvement into aircraft control and checklist utilization tasks during safety critical phases of flight may produce positive results required for their future successful professional careers.

Methodology

The methodology of this research is based on quantitative analysis of flight crewmembers' activity data (maintaining of the flight path parameters and application of CRM skills) influenced by variable levels of checklist utilization during several manually controlled instrument training ILS approach flights in similar conditions simulated in a Piper Seneca IV airplane FTD.

Grounding of Simulated Flight Conditions

Following factors were considered for grounding of simulated flight conditions used for the research. *Distractions of pilots' attention* from perceiving the aircraft flight path parameters, from proper communication between crewmembers, and from applying control inputs may significantly reduce the quality of the flight path parameters, as well as reduce application of the crewmembers' CRM skills.

Although the distractions may occur during every phase of any fight (automatic, manual, visual, instrument), the worst negative influence of distractions may be expected in the course of final approach in manually controlled instrument flight. Thus, the final approach phase of the flight was chosen for the experimental exercises.

A well-described reference flight path line is needed to quantitatively assess aircraft flight path deviations caused by the PF distractions from controlling the aircraft. An ILS approach procedure may produce such a reference line.

Thus, the manual control of flight path in the course of the ILS approach was chosen as an indicator of crew performance in controlling the aircraft while this performance was interfered with distractions caused by the checklist utilization.

The Study Data Collection

The process of data collection occurred over the line oriented flight crew simulation course student flight training required by the WMU program.

The experimental FTD portion of the research was performed by subjects during simulated flight training exercises within a part of the final approach flight path located between the ILS glide slope (GS) capture point and the Decision Height (DH) point. *Flight crew performance data* needed for assessments of the flight crew activity interfered with checklist utilization were collected within this part of the flight path. Circled numbers in Figure 1 show locations of the Piper Seneca IV airplane checklist sub-sections utilization actions performed by crews during flight exercises. The bold dash line shows the required flight path that the flight crewmembers had to maintain.

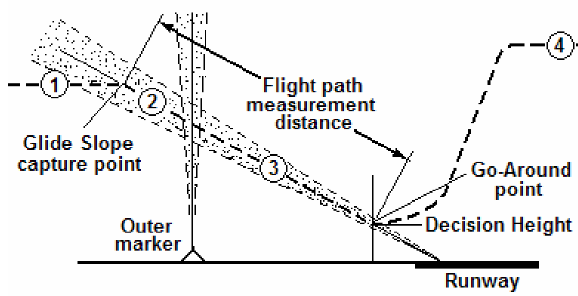


Figure 1. *The FTD experimental exercise diagram*

Every exercise began before the "airplane" (simulated by the FDT) reached the GS (point 1). The airplane wing flaps were in 10 degrees deployed position, as recommended by the simulated airplane Information Manual (Piper Aircraft, 1995). By this moment of the flight the flight crew completed approach briefing, and then they used the "down to the line" section of the "Pre-Landing" checklist section.

After the GS capture and lowering landing gear the "below the line" sub-section of the "Pre-Landing" checklist section was used (point 2). The "Landing" checklist section was performed after the "aircraft" passed the outer marker and was established on the GS (point 3). The cloud base imitated in the FTD was below the landing minimum, and the crew initiated the go-around (GA) maneuver at the DH. After actions required by the GA procedure had been performed, pilots completed the "After Takeoff / Go-Around" checklist section (point 4), the FTD was "frozen" by the instructor, and the exercise was terminated.

The actual flight path parameters in a graphic form and instructor's evaluation of the crewmembers' CRM skills in a numerical form were recorded for every exercise. The researcher then statistically processed the collected data and created a document representing the data in a quantitative form convenient for the research results analysis.

Flight Crew Activity Scenarios

Two experimental flight crew activity scenarios were designed to assess the checklist utilization influence on the crew performance. The "challenge-do-verify" crew activity of checklist utilization was used for the whole FTD experiment. The first scenario was used for collecting crew performance data influenced by variable degrees of checklist utilization. The second scenario represented a control group of research subjects created to compensate for pilot training effect during the series of flight exercises. Every research subject crewmember performed three ILS approach exercises.

In the first scenario the degree of crewmembers' involvement into the checklist utilization changed from a minimal to full levels.

During the first ILS approach exercise crewmembers did not pronounce the checklist callouts and responds, and the PNF used the checklist silently only as a reference to avoid missing of required crew actions in the cockpit. The only task of the PF in the first exercise was to manually control the airplane flight path using the ILS indications. The PNF was instructed to pay maximum of their attention to aircraft flight path parameters monitoring, and to inform the PF about the parameters deviations.

During the second exercise the PNF read aloud the checklist callouts and responds, and performed all required actions. Because of this additional workload the PNF partially diverted their attention from the flight path parameters monitoring and from timely informing the PF about the parameters deviations. *During the third approach* the PNF read the checklist callouts, and both pilots performed actions in areas of their responsibility in the cockpit, and pronounced responds confirming completed actions. These actions distracted both of the pilots from the flight path control.

In the second scenario performed by subjects of the control group, to compensate for the effect of student pilots' flight path control skill increase in the result of successive identical flight exercises, the PF and the PNF were instructed to utilize the checklist in full during all three exercises. The control crew PNF read the checklist item callouts, and both of the control crew pilots performed all required checklist actions in accordance with their roles in the flight control process and areas of responsibility in the cockpit.

Participating Subjects

Subjects participated in the research were the Western Michigan University College of Aviation pilot students.

The instrument flight proficiency of student pilot subjects participating in the experiment was approximately equal. Most of subjects did not operate the simulated type of the aircraft before the experiment. To have the subjects familiar with the FTD cockpit layout the FTD instructors provided an introductory session for all crews before the experiment.

Instruments for the Study

The instruments for this study included:

- a flight training device (FTD) of the Piper Seneca IV airplane controlled by two pilots;
- an ILS approach procedure of the Battle Creek, Michigan, regional airport (KBTL) with a portion of flight path including the final approach and the go-around phases;
- standard operational procedures required by the simulated airplane Information Manual;
- "normal" checklists for the final approach and go-around procedures;
- automatic record of flight path parameters maintained by subjects in the course of performing the exercise procedures;
- FTD instructors' evaluations of student pilot CRM skills.

Quantitative Assessment of Crew Performance

Two quantitative criteria were used in FTD training exercises to assess crewmembers' performance influenced by the checklist induced distractions: a flight path parameter maintaining criterion "Δ" calculated from actual flight path parameter values, and an empiric numerical evaluations of crew CRM skills given by the FTD instructor.

A degree of the flight path curvature expressed through a standard deviation was chosen as a crew flight path parameter maintaining criterion. The criterion was calculated under a formula (1) used by the researcher in one of his previous studies (Risukhin 1988):

$$\Delta = \sqrt{\frac{(\sum_{i=1}^{n} h_i^2 - a\sum_{i=1}^{n} h_i - b\sum_{i=1}^{n} l_i h_i)}{n}} \quad (1),$$ where

Δ - the flight path parameters maintaining criterion,

h_i (height) and l_i (distance) – current values of the actual flight path points measured relatively to the intended landing runway,

i = (1, 2, … , n) - quantity of h_i and l_i pairs;

$$a = \frac{\sum_{i=1}^{n} h_i - b\sum_{i=1}^{n} l_i}{n}, \qquad b = \frac{\sum_{i=1}^{n} l_i h_i - \frac{\sum_{i=1}^{n} l_i \sum_{i=1}^{n} h_i}{n}}{\sum_{i=1}^{n} l_i^2 - \frac{(\sum_{i=1}^{n} l_i)^2}{n}}.$$

The crew CRM skills numerical evaluations range: 1 (poor), 2 (satisfactory), 3 (good), and 4 (very good).

Preliminary Results

Preliminary FTD experiments were performed in accordance with the described research methodology.

Differences were noted between student pilot crew performance assessments obtained in two instrument flight training scenarios with various degrees of checklist-induced crew distractions.

Notwithstanding a seeming increase of the pilots' attention share paid to the checklist in the first scenario from the first exercise through the third one, preliminary results have shown an improvement of flight path parameters maintaining in the third exercise of both scenarios. Two possible causes of this fact may be considered:

- increase of the PF instrument control proficiency due to repeating of the exercises;
- improvement of coordination between PNF and PF in flight path maintaining and checklist utilization.

Further experimental FTD exercises are needed to obtain statistical data sufficient for analysis of correlations between checklist utilization way as an independent variable, and assessments of flight crew performance in aircraft flight path parameter maintaining and CRM skills as dependent variables.

Discussion and Conclusions

The research of checklist induced flight crew distraction is based on analysis of flight crew performance criteria obtained in FTD pilot training exercises.

Two scenarios of checklist utilization as an independent variable during the series of training exercises were designed for the research: a gradual increase of crew distraction caused by the checklist utilization, and a continuously high involvement of both crewmembers into checklist utilization.

Measurements of flight path parameter deviations from required values, and instructors' evaluations of the crewmembers' CRM skills are used for the crew performance assessment.

Accumulation of the research data is needed to allow quantitative comparison of student pilots training results in flight path parameter maintaining, and in their coordination and interaction, interfered by various techniques of the checklist utilization.

Gradual increase of the PF involvement into the checklist utilization, optimization of the PNF attention distribution between flight path parameter monitoring and checklist utilization, and improvement of PNF and PF interactions in the course of several identical training exercises may help in increase of student pilot instrument flight control proficiency.

Optimization of student crewmembers interaction and callout / response / monitoring / control actions coordination between the student pilots in the process of instrument flight training with checklist utilization may improve developing of their skills in overcoming distractions from one of the most safety critical tasks - immediate controlling of the aircraft flight path.

The proposed methodology may be applied for research and improvements of various instrument flight training programs, including manual and automated non-precision approach training.

Many of aircraft accidents were caused by insufficient instrument flight proficiency of their pilots. A significant reduction of checklist induced flight crew distraction may be reflected in pilot training programs as well as in operation procedure design. These steps may help to increase reliability of crewmember flight path control performance.

References

Degani, A. & Wiener, E. (1990). Human factors of flight-deck checklists: the normal checklist. NASA Contract Report 177549. Ames Research Center, Moffett Field, California.

James, William (1890**).** The Principles of Psychology. Retrieved on 01/30/2005 from http://psychclassics.yorku.ca/James/Principles/prin11.htm.

O'Hare, D. & Roscoe, S., with contributions by Vette, G. & Young, M. (1990). Flightdeck performance: the human factor. Iowa State University Press, Ames.

Risukhin, Vladimir (1988). An analysis of factors influencing crew performance, and development of a methodology for non-scheduled flight operations planning. Academy of Civil Aviation, Leningrad, Russia (in Russian).

The New Piper Aircraft, Inc. (1995). Seneca IV PA-34-220T Information Manual. Manual part number 761-872.

U.S. Department of Transportation Federal Aviation Administration (1995). Human performance considerations in the use and design of aircraft checklists.

U.S. Department of Transportation Federal Aviation Administration (2000). Standard operating procedures for flight deck crewmembers. Advisory Circular AC No: 120-71.

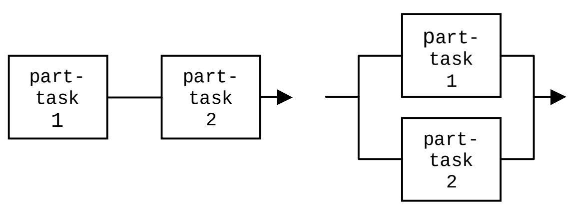
TRANSFER BETWEEN TRAINING OF PART-TASKS IN COMPLEX SKILL TRAINING – MODEL DEVELOPMENT AND EXPERIMENT

Jan Joris Roessingh
National Aerospace Laboratory NLR
Amsterdam, the Netherlands
Astrid M.L. Kappers & **Jan J. Koenderink**,
Helmholtz Instituut, Universiteit Utrecht, the Netherlands.

One of the most common instruction-strategies for training complex skills is part-training. In this paper a model is developed for the optimisation of schedules for part-training, the ‘optimal transfer model’. This model is based on individual learning, but may be generalised to groups of trainees. It is based on the idea that if there is functional skill transfer from part-training to whole-task performance, then there must be a training schedule that yields optimal results. In this context, an optimal training schedule is one in which part-training lasts as long as is necessary to ensure the best possible performance with the whole-task at the end of the training. To prove that an optimal training schedule does in fact exist, an experiment was conducted in which different groups of trainees received sixteen hours of training under different part-training regimes to learn a complex vehicle control task. The individual learning curves of all trainees were measured. Application of the optimal transfer model to the learning curves allowed determining the optimal part-task schedule. Applications of the model to practical training situations are discussed.

Introduction

Dividing the whole task into part-tasks



a. segmentation b. fractionation

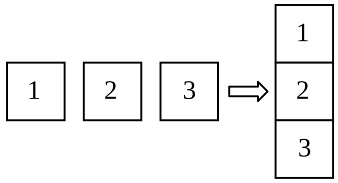
Figure 1. *Serial and parallel execution of part-tasks*

Part-training (or part-task training) has been defined as the training of a number of separate components (part-tasks) as the precursor of practising the whole task. The basic principles of part-training are twofold: (1) the separation of the whole task into part-tasks and (2) the scheme for integration of the part-tasks during training. According to Wightman & Lintern (1985) the whole task can be divided into part-tasks in three basic ways: (1) segmentation, (2) fractionation and (3) simplification.

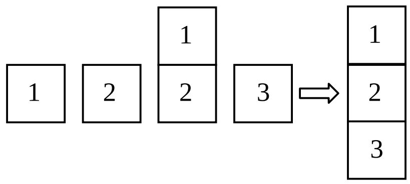
When the task is divided along spatial or temporal dimensions, the division is called <u>segmentation</u> (figure 1a). This method applies when task components have a clear beginning and end in space or time, i.e. when different task components are executed serially in the whole task. An example from aviation is the handling of the Control and Display Unit (CDU), which can be considered to be a segment of the whole flight task. Other examples are in-line parking of a car or joining the traffic when you drive onto the motorway, both of which can be considered as segments of the whole driving task. <u>Fractionation</u> (figure 1b) applies when different task-components can be executed in parallel in the whole task. For example, the control of pitch, roll and yaw channels in co-ordinated aircraft manoeuvres can be considered as separate fractions, or checking the rear-view mirror in a car manoeuvre can be considered as a fraction. Finally, <u>simplification</u> applies when part-tasks are the result of the modification of features of the whole task. An example is the reduction that is made in the number of aircraft per unit time entering the controlled airspace in a simulated air traffic control task.

Re-integration of part-tasks during training

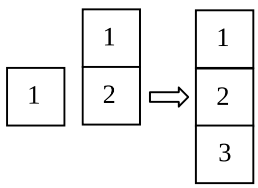
Reconstruction of the whole task from its part-tasks in the course of the training can proceed according to different schemes (figure 2). The basic schemes for task integration are (1) pure part-training, (2) progressive part-training and (3) cumulative part-training. In pure part-training, all part-tasks will be practised separately before the whole task is tackled. In progressive part-training each part-task will first be practised separately and subsequently together with the preceding part-tasks. Finally, in cumulative part-training, only the first part-task will be practised separately, subsequently a second part-task is added, etc.



a. pure part-training



b. progressive part-training



c. cumulative part-training

Figure 2. *Integration of part-tasks*

Benefits of part-task training

Part-training can be applied for two different reasons. The first reason is that part-training can often be carried out with relatively simple training media. If training with these 'part-task trainers' replaces training with more sophisticated media (e.g. full mission simulators), training costs can be reduced. The second reason is that part-task training can be more efficient, i.e. it speeds up the learning process and thus saves training time. It is generally assumed that the increased efficiency of part-training will occur only in the initial phase of the learning process and will be particularly beneficial if the task is highly complex and if trainees are of lower than average ability. In a case where the task is too complex for the trainee, exposure solely to the whole task may even prevent the learning process from starting.

However, reviews of training research (Wightman & Lintern, 1985, Teague, Gittleman & Park, 1994) indicate that in the majority of cases part-training is less efficient than whole task training. Whole-task training thus is the preferred method if the task is sufficiently simple and can be reasonably approximated by the trainee. Only when the whole task is dangerous or highly complex and can be easily divided into part-tasks is part-training the better choice. Teague and colleagues (1994) argued that with regard to recall and recognition context-dependent methods are favoured over context-independent methods. However, if the acquired knowledge and skills have to be selectively applied in a variety of situations, context-independent presentation methods are recommended.

A model for part-task efficiency with one part-task

In this study 'speed-tasks' (or 'speed-based tasks') are investigated. These are tasks that allow the skill-level to be measured principally by the trainees' speed of performance, once completion of the task can be taken for granted. Crossman (1959), in his classical study of skill acquisition in cigar production, used the term speed-task. The production of one cigar, with specified quality, could be measured merely in terms of production time ('trial time') or its reciprocal: production rate, i.e. number of completed products per unit time.

Throughout this study, the term 'trial time' rather than 'response time' is used, to indicate the time needed to successfully complete a trial on a speed-task. After all, a complex task encountered in the real world usually requires a series of responses rather than a single response and hence 'response time' is an inappropriate term.

Learning models for speed-tasks have sometimes been expressed as *rate* models (e.g. Restle & Greeno, 1970, Mazur & Hastie, 1978, Gallistel & Gibbon, 2000). These models assume that learning is based on the temporal intervals between events and the reciprocals of these intervals, the rates at which events occur. In this study we define rate λ as the reciprocal of the time interval between subsequent successful trial completions, i.e. as the reciprocal of trial time. Series of such trial times $\{T_1, .., T_n\}$ are measured to investigate the changes that occur in individual skill level during practising a task. Hence, the rate λ at which a trainee completes the n^{th} trial on a task is by definition:

$$\lambda = \frac{1}{T_n}. \qquad (1)$$

It is assumed that this rate λ increases linearly with training time (i.e., cumulative trial time: $t = T_1 + T_2 +.. + T_n$). Following this assumption, a functional expression for the expected rate $E[\lambda(t)]$ at which subsequent trials on a whole task will be completed is:

$$E\left[\lambda(t)\right] = a \cdot t\,. \quad (2)$$

According to equation (2) the expected rate increases in proportion to training time t. The parameter a is a proportionality constant. It represents the increase in rate λ per unit training time t. Since the dimension of t is s, and λ has the dimension trials/s, parameter a must have the dimension trials/s^2. If we were to plot the *learning curve* $\lambda(t)$ against t, the parameter a would represent the tangent (slope) of the best fitting straight line through this curve.

Now, assume that before practising the whole task, the trainee has practised with a part-task during a period of practice time t_*. If there is transfer between the part-task and the whole task, part-practice would cause the slope a of the learning curve to change with a quantity a_* and would cause a constant bias λ_* in whole-task performance. In accordance with equation (2), a functional expression for the rate at which trials are completed during subsequent whole-task practice then becomes:

$$E\left[\lambda(t)\right] = \lambda_* + (a + a_*) \cdot (t - t_*)\,. \quad (3)$$

It is further assumed that the change in slope a_* is a linear function of part-task practice time t_*[1]. This gives the 'transfer function':

$$a_* = b \cdot t_*\,, \quad (4)$$

in which the parameter b is a proportionality constant, representing the constant increase in learning speed on the whole-task per unit practice time t_* with the part-task. Parameter b has dimension trials/s^3 (since a_* has dimension trials/s^2 and t_* has dimension s).

When we substitute equation (4) into equation (3), we obtain:

$$E\left[\lambda(t)\right] = \lambda_* + a \cdot t + b \cdot t \cdot t_* - b \cdot t_*^2 - a \cdot t_*\,. \quad (5)$$

Optimal transfer of part-training occurs when practice with the part-task produces the maximum skill-level on the whole-task at the end of the training. Thus, when total training time has a limited duration in which both the part-task and the whole task must be practised, the 'logistics' question is: How can one achieve the highest rate λ on the whole task at the end of the training? In other words, we must find the value for part-task practice time t_* that maximises the expected rate $E[\lambda(t)]$ given a fixed (limited) training time t.

A functional expression for the value of t_* that maximises $E[\lambda(t)]$ can be found by taking the first derivative of $E[\lambda(t)]$ in equation (5) with respect to t_* and setting this derivative to zero[2], which gives:

$$b \cdot t - 2 \cdot b \cdot t_* - a = 0\,. \quad (6)$$

Hence, the optimal practice time t_* with the part-task is:

$$t_{*opt} = \tfrac{1}{2}\left(t - \frac{a}{b}\right). \quad (7)$$

Note that with any combination of positive values for the constants a and b, the optimal training time t_{*opt} with the part-task is less than fifty per cent of the total training time t. The solution of equation (7) can be substituted into equation (5) to give the corresponding optimal performance:

$$E\left[\lambda(t)\right]_{opt} = \lambda_* + \tfrac{1}{4}\tfrac{a^2}{b} + \tfrac{1}{2} \cdot a \cdot t + \tfrac{1}{4} b \cdot t^2\,. \quad (8)$$

The optimal training time t_{*opt} with the part-task and the corresponding optimal performance can be calculated once values for the free parameters a, b and λ_* are known (or rather, when these parameters can be estimated from data collected during training).

More elaborate models for transfer can be obtained when formulations like those of equation (5) are based on more general models for the learning curve, rather than on a simple linear function (2). Moreover, the model could be based on more general transfer functions than the simple linear function of equation (4), and the model could be further generalised for a multiple part-task scheme, rather than a simple scheme with one part-task only. However, for current purposes, and in the absence of evidence needed for a more elaborate model, the simple model of equation (5) will be investigated empirically.

---

[1] For ease of exposition we silently assume that λ_* is an arbitrary constant, independent of t_*.

[2] To find a maximum it is also necessary for the second derivative of $E[\lambda(t)]$ to be negative, which, in this case, requires that parameter $b>0$.

Method

Tasks

In the two different versions of the Space Fortress-game (SF-game) used in this research and described below, the display contains a rotating fortress in the centre and a manoeuvrable spaceship, which has a starting position in the lower right corner of the display. The trainee controls the spaceship's flight with a joystick. The trajectory of flight can be controlled by rotating the ship and applying thrust (which causes the ship to accelerate). The ship continues to fly in the direction in which it is pointing, unless it is rotated and thrust is applied. This 'control law' significantly contributes to the complexity of the task, since novice trainees do not learn the law intuitively or easily.

The part-task contains only a subset of the game elements of the full SF-game (Mane & Donchin, 1989). This part-task was used previously by Frederiksen & White (1989). The trainee controls the spaceship's flight with a joystick and fires missiles from the ship by pressing a fire button on top of the joystick. The trainee's task is to attack the fortress by hitting it ten times with a missile, at intervals of at least 250 ms, before destroying it with a burst of two shots (fired at an interval of less than 250 ms).

The fortress defends itself against the ship. It does this by rotating to face the ship and then tracking the ship's movements while firing shells at it. When the ship is hit for the fourth time by a shell from the fortress, it is returned to its starting position. When this happens, the shot counter, which counts the hits scored against the fortress, is set to zero. A trial on the task finishes as soon as the fortress is destroyed.

The whole task is the full SF-game. The fortress is protected by moving 'mines' which emerge on the display periodically. These mines chase the ship. Unless the trainee takes action, these mines will hit the ship. Moreover, when a mine is present on the display, missiles fired at the fortress have no effect. Thus, the mine has to be eliminated by a missile immediately. However, if the trainee fails to hit the mine within 10 seconds, the mine disappears from the screen automatically. The interval between the disappearance of one mine and the appearance of the next is four seconds, during which time the trainee can fire at the fortress. When the ship is hit for the fourth time by either a mine or a shell from the fortress, the ship is returned to its starting position and its shot counter is set to zero. As in the case of a part-task, a trial on the task finishes as soon as the fortress is destroyed. What makes the whole task even more complicated is that the trainee has to distinguish between two types of mines, and react accordingly. The more difficult mine can be identified by a letter that appears in the information panel at the bottom of the screen (prior to each five-minute block of play, the trainee is presented with a new set of three letters that are used to identify 'difficult' mines). The appearance of a difficult mine requires the trainee to press the right ('identification') button on the mouse twice of an interval of 250-400 ms before the mine can be destroyed by a missile. The 'easy' mine can simply be destroyed by hitting it with a missile without pressing the identification button. However, if a trainee mistakenly presses the identification button and the mine is not an 'easy' one, the mine becomes invulnerable to missiles; then it cannot be eliminated and will either hit the ship or automatically disappear after 10 seconds. Since missiles fired at the fortress have no effect when a mine is present, the trainee can choose whether to avoid the invulnerable mine and wait for it to disappear or let it damage the ship. Another complication in this task is that the supply of missiles is limited, and the stock has to be monitored in the information panel at the bottom of the screen. An extra supply can be obtained by using 'resource opportunities'. The availability of these opportunities are indicated by a random sequence of symbols (&, #, $, %, !, etc.) which appear in the centre of the display (beneath the fortress). When the $ symbol appears for the second time in a row, the trainee can get extra missiles by clicking the middle button of the mouse. As with the part-task, a trial finishes as soon as the fortress is destroyed.

Trainees

Twelve male university undergraduates aged between 20 and 23, with normal vision, participated in the study. Trainees were recruited via an advertisement in the University magazine of Utrecht University. In total 36 trainees were selected from a larger group of 51 candidates by means of the Aiming Screening Task (AST), a task that is known to be a reasonable predictor for training success on this task (see Foss, Fabiani, Mane & Donchin, 1989). An AST-score of 740 points was the minimum score required for participation in the study. As the current study is part of a larger training study, the sixteen trainees in the current study are a balanced subset of the full set of 36 trainees who participated in the larger study. The subset has the same average AST-score (870 points) as the full set, and each trainee with an above-average AST-score is paired with a trainee with a below-average score. None of the trainees reported playing video games for more than 4 hours per week.Trainees were paid 30 Euro per day plus a bonus of 68 Euro upon completion of the experiment.

Procedure

Trainees were assigned either to the whole task group or to the part-task group. The assignment was balanced between the two groups on the basis of the AST-score achieved. The six trainees assigned to the whole-task group practised with only the whole task (the full Space Fortress game) and received no previous practice training on a different task. The six trainees assigned to the part-task group first practised with the part-task and thereafter practised with the whole task.

Two trainees of the part-task group transferred to whole-task practice after $t_* \approx 6000$ s (100 minutes), two trainees transferred after $t_* \approx 12000$ s (200 minutes) and two trainees transferred after $t_* \approx 36000$ s (600 minutes ~ 10 hours).

Total practice time (time-on-task) was 16 hours in total for all trainees in both groups. To this end, eight training days over a five-week period were scheduled for each trainee. During a training day, the trainee would complete three training sessions consisting of eight blocks of five minutes each, separated by two breaks of twenty minutes. The effective time-on-task was thus forty minutes per session and 120 minutes per day. Trainees were allowed to take one-minute breaks between five-minute blocks. The data collected with the six trainees in the whole-task group have been published previously in Roessingh, Kappers and Koenderink (2002). The data collected with the six trainees in the part-task groups have not been previously published.

Software and equipment

The experiment room contained individual computer stations in separate cubicles. Each computer station was equipped with a PC and a joystick of type FlightStick (CH-products). The joysticks were modified so that they could be connected to an A/D converter card (DataTranslation) in the PCs. The fire-button on the joystick and the three other response buttons were connected to a timer card in the PC. A camera system was installed in the cubicles to control the course of the experiment.

The original SF software was made available by the Dept. of Psychology, University of Illinois at Urbana-Champaign. To facilitate Task 1 and Task 2, the software was modified to remove the specified components of the full SF-game. The software was also modified to record additional parameters, in particular total time-on-task and trial-times, with a timing accuracy of 50 milliseconds.

Further training materials

After screening and well before the start of the experiment, the trainees received the instruction booklet for the SF game by mail at their home address. This instruction booklet specified the rules of the game and explained how to control of the space ship. No reference was made to specific tactics or strategies. The trainees were instructed to study the booklet carefully before the experiment began.

Results

Whole-task learning curves

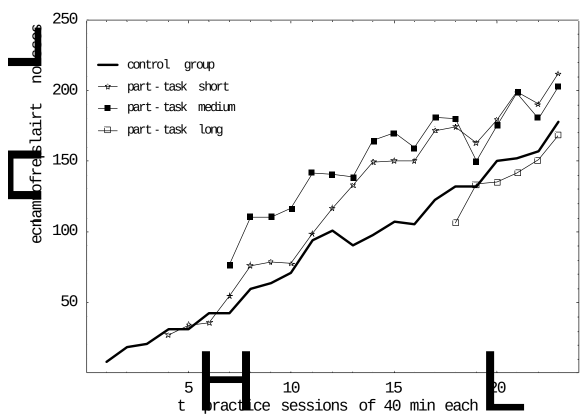


Figure 3. *Average learning curves with the whole task for the control group (6 trainees, long solid line) and for each part-task condition (2 trainees per condition, after short, medium and long training on the part-task).*

Figure 3 represents learning data of all trainees during practice on the whole task, the full SF-game. The horizontal time axis denotes practice time in units of 40 minutes each (each of the 24 practice-sessions took 40 minutes). The vertical axis denotes task performance (the number of fortresses destroyed). Hence, each data point is the number of fortresses destroyed in a particular session.

The thick solid line is the learning curve for the whole-task (control) group. Performance per session has been averaged over the six trainees in this group. The line with the star-symbols is the average learning curve of the two trainees who transferred to the whole task after ~100 minutes of part practice. The somewhat shorter line with the filled box-symbols is the average learning curve of the two trainees who transferred after ~200 minutes of part practice, and the shortest line is the average learning curve of the two trainees that transferred after ~600 minutes of part practice.

The relative location of the learning curves of figure 3 suggests that practice with the part-task generally had a positive effect on whole-task performance, particularly for the trainees who transferred after 100 and 200 minutes. Moreover, the curves for these trainees suggest that the latter made more efficient use of training time.

The "linear rate assumption"

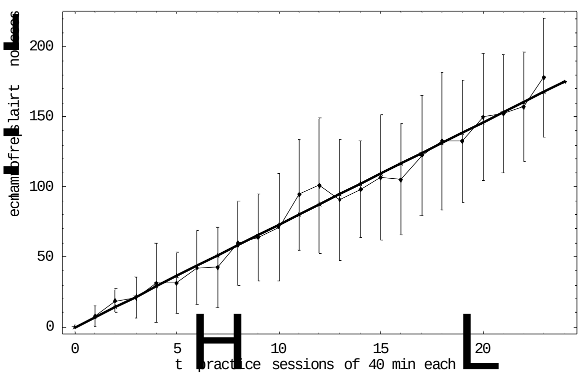


Figure 4: *Learning curve for the control group with the best fitting model $\lambda=a\cdot t$ (the straight solid line). Error bars indicate the standard deviation in performance over the six trainees.*

The model presented for part-task transfer is based on the assumption that the rate at which trials on a speed-based task are completed will increase linearly with practice time; this was expressed by the formula $E[\lambda(t)]=a\cdot t$, i.e. equation (2).

To check whether this assumption is correct, the learning curve of the control group is represented separately in figure 4. As with figure 3, performance λ is plotted against time t. The error bars represent the standard deviation in the performance score λ of the six trainees. The percentage of variance accounted for by the linear model of equation (2) is 90 per cent (R^2=0.90). The slope a of this model can be estimated from the data, which slope is 7.3 trials/session·session ($1.3\cdot10^{-6}$ trials/s^2). The null-hypothesis for the straight-line fit, which states that the slope a equals zero, has to be rejected ($T(143)=34.7$, p 10^{-6}).

The "linear transfer assumption"

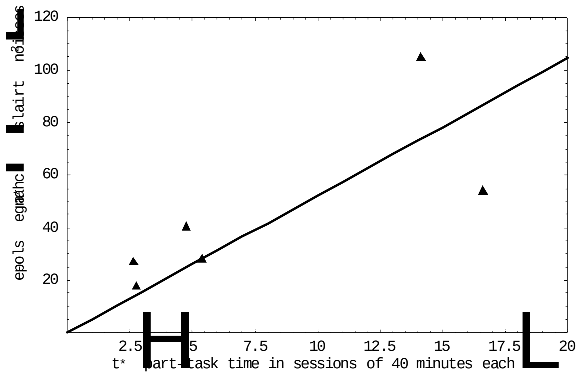


Figure 5: *Change in slope (a_*) of the whole-task learning curve as a function of practice time t_* with the part-task. Each data point corresponds to the calculated slope change a_* of a trainee transferring to the whole task after t_* minutes. The fit of the model $a_*=b\cdot t_*$ is based on six data points (six trainees).*

The other basic assumption of the model for part-task transfer concerns the linearity of transfer from the part-task, the assumption being that the increase in the tangent (slope) of the whole-task learning curve is proportional to practice-time with the part-task.

In the preceding section, the assumption that the learning curve during whole-task practice is a linear function was considered. This provided the basic learning curve equation $\lambda=a\cdot t$. It was assumed additionally that part-task practice with duration t_* causes the slope a to change with a fraction a_*. More specifically, it was assumed that the slope-change a_* is linear with practice time t_* on the part-task, such that the slope-change a_* satisfies the equation $a_*=b\cdot t_*$, cf. equation (4).

In figure 5, the slope-change a_* is plotted for the six trainees who practised with the part-task. This slope-change a_* for each of the six trainees has been calculated as the difference between the slope a of each of these trainees and the average slope a of the trainees in the control group (figure 4). The horizontal axis plots the number of minutes t_* that the trainees spent practising the part-task. The solid line in figure 5 is the best fitting model $a_*=b\cdot t_*$ with ordinary least squares. The constant model parameter b is estimated to be at 5.2 trials/session^3 (~$3.6\cdot10^{-10}$ trials/sec^3). The linear model accounts for 86 per cent of the variance in these six data points (R^2 = 0.86) and b is significantly different from zero ($T(5)=5.5$, $p<0.003$). Although the fit is based on six data points only, the linear transfer assumption does not seem unreasonable.

Optimal training time for the part-task

Once an estimate of the constants *a* and *b* is obtained, we can use equation (7) to determine the optimal training time $t_{*\text{opt}}$ with the part-task. In the preceding sections we estimated the slope *a* of the whole-task learning curve to be 7.3 trials/session^2 and we estimated the constant *b* to be 5.2 trials/session^3.

Since total training time was fixed at t = 16 hrs = 24 sessions of 40 minutes in this training experiment, we can calculate $t_{*\text{opt}}$ with equation (7) as:

$$t_{*opt} = \tfrac{1}{2}\cdot 24 - \tfrac{1}{2}\cdot\frac{7.3}{5.2} = 11.3\,sessions\,.$$

Expressed as a percentage of total training time *t*, optimal training time with the part-task is:

$$\frac{t_{*opt}}{t} = \frac{11.3}{24} \quad 47\%\,.$$

Discussion

In this study, two groups of trainees received experimental training with a complex task: the Space Fortress game (SF). We used SF because this game is a representative skill trainer for complex tasks encountered in the real world, such as flying an aircraft. The statement that SF is representative for this type of tasks is supported by field studies at flight schools where SF has been used in flight training. Examples are research with the Israeli Air Force by Gopher, Weil & Bareket (1992, 1994), with the US Army by Hart & Battiste (1992) and with the US Air Force by Vidulich, McCoy & Crabtree (1995).

In the experiment described in this research, the control group received training with the full SF game only. The experimental group first received part-task training with a simpler version of the game, from which the cognitive components were removed such that the emphasis was on manual control. We analysed the learning curves of the trainees in both groups in order to verify a quantitative model for skill transfer. Skill transfer (transfer of training) deals with the degree to which learning a target task (in this case, the full SF-game) is facilitated by the prior learning of another task (in this case the part-task, the simpler version of the SF-game).

Testing the two assumptions of the model

The model presented in this paper is based on two simplifying assumptions. The first is the “linear rate assumption”, which states that individual skill-level on a speed-based task (measured as a performance rate) increases linearly with practice time. The second is the “linear transfer assumption”, which states that there is a linear relationship between the amount of prior practice with the part-task (measured in units of practice time) and the slope of the individual learning curve measured on the target task.

The first assumption, the linear rate assumption, sounds odd, since people tend to think that learning is initially fast and then gradually slows down towards an asymptote; hence learning curves are usually considered to be non-linear. Nevertheless, the present data show that for complex speed-tasks, i.e. tasks with no speed-accuracy trade-off and ample opportunity for speed-improvement, the linear rate model is an approximate description of the data. It should be noted that alternative, more complex, models, such as higher-order linear models or non-linear models have not been tested. In future research, plausible alternatives for the linear rate model could be developed and tested against it. At the present time, the linear rate model seems a reasonable approximation for 16 hours training with the full SF-game, presumably since this task is sufficiently complex and interesting to guarantee a much longer skill acquisition process until the asymptote is reached. It is not within the scope of this paper to present a theoretical justification for the linear rate assumption. However, such theory can be found in Roessingh et al (2002).

The rationale of the second assumption, the linear transfer assumption, and its plausibility, are similar to the rationality and plausibility of the first assumption. The interpretation of a time-linear increase in performance rate as a result of repeatedly practising a task, is that during practice there is linear transfer from one time-unit to the next. Thus, given the plausibility of time-linear transfer within a single task, a similar transfer characteristic between different tasks should be equally plausible; this provides us with the basis for the linear transfer assumption. Since this assumption could only be analysed and verified on the basis of the learning curves of six trainees in the experimental group, more research is needed to further test and understand skill transfer in the acquisition of complex skills.

Predictions of the linear transfer model

We argued that, on the basis of the model presented, the optimal training schedule can be predicted, given the credibility of its assumptions and appropriate estimates for the parameters *a* and *b*. In the results section we provided the optimal schedule for the training that we used in the experiment. But even in the absence of such appropriate estimates, the model makes interesting predictions. For example, it should be noted that, for any positive *a* and *b*, equation (7) implies the following inequality:

$$\frac{t_{*opt}}{t} < 50\%, \qquad (9)$$

such that optimal training time $t_{*\mathrm{opt}}$ with one part-task is always less than fifty per cent of the total training time *t*.

Note that a negative value for the slope *a* would indicate a decreasing learning curve as a result of practice, whereas a negative value for the constant *b* would be a matter of negative transfer from the part-task. In these (dubious) cases, the model presented for "optimal" transfer, based on determining optimal performance by solving from equation (5):

$$\frac{dE\left[\lambda(t)\right]}{dt_*} = 0, \qquad (10)$$

would identify training schedules for minimal performance rather than maximal performance. Hence, situations in which either *a* or *b* is negative should therefore be avoided. The case in which both learning curve slope *a* and transfer *b* are negative seems to be entirely theoretical.

Applications

Since the linear transfer model can be used to predict optimal training schedules, it can be applied for the professional training of complex skills. The present model is applicable to the acquisition of speed-skills in training situations with one part-task. An example is the training that pilots receive on the ground, with a part-task trainer, a procedure trainer or a simulator, to learn a set of instrument procedures. After the training on the ground, training in the real aircraft is provided. The model can be used to decide on the ideal ratio between time spent training on the ground and the time spent training in the air.

Examples of speed-skills suitable for part-training can be found in a wide range of domains: air traffic control, military aviation and industrial manufacturing, to name but a few.

It seems fairly straightforward to generalise the model to schemes with multiple part-tasks, rather than restrict it to a simple scheme with one part-task only. Moreover, the model could also be generalised to accuracy-based tasks, rather than speed-tasks only. With these generalisations the model can potentially be applied in many situations in which complex skills must be acquired and an appropriate training time schedule has to be worked out. Obviously, a practical and useful version of the model would also take into account the relative cost per unit time of part-task training and whole-task training.

Conclusion

The results demonstrate that a simple two-parameter model (the 'linear transfer model', which is based on two assumptions about the nature of learning and transfer) can be used to predict the optimal training time schedule in part-task training. An interesting prediction of the model is that, in training with only one part-task, more than fifty per cent of the total training time should be devoted to practice with the whole task in order to maximise performance. This prediction does not depend on the precise parameter values in the model. However, when reliable parameter values can be obtained, more accurate predictions can be made, as was demonstrated with the data from the training experiment. The linear transfer model can be applied in training situations where trainees need to acquire speed-skills, for example in military aviation.

Acknowledgements

We thank TNO Human Factors Research Laboratory, particularly Dr. John van Rooij (now with the Netherlands MoD), for enabling the data collection. We are grateful to the twelve students who took part in the experiment.

References

Crossman, E.R.F.W. (1959). A theory of the acquisition of speed-skill. *Ergonomics*, Vol. 2, 153-166.

Foss, M.A., Fabiani, M., Mané, A.M. & Donchin, E. (1989). Unsupervised practice; the performance of the control group. *Acta Psychologica*, Vol. 71, 23-51.

Frederiksen, J.R. & White, B.Y. (1989). An approach to training based upon principled task decomposition. *Acta Psychologica*, Vol. 71, 89-146.

Gallistel, C.R. & Gibbon, J. (2000). Time, rate and conditioning. *Psychological Review*, Vol. 107, No. 2, 289-344.

Gopher, D., Weil, M. & Bareket, T. (1992). The transfer of skill from a computer game trainer to actual flight. In: *Proceedings of the Human Factors Society*, 36th annual meeting.

Gopher, D., Weil, M. & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. *Human Factors*, 36(3), 387-405.

Hart, S.G. & Battiste, V. (1992). Field test of video game trainer. In: *Proceedings of the Human Factors Society*, 36th Annual Meeting.

Mané, A. & Donchin, E. (1989). The Space Fortress Game. *Acta Psychologica*, Vol. 71, 17-22.

Mazur, J.E. & Hastie, R. (1978). Learning as accumulation: a reexamination of the learning curve. *Psychological Bulletin*, Vol. 85, 6, 1256-1274.

Restle, F. & Greeno, J.G. (1970). *Introduction to mathematical psychology*. Reading, Mass.: Addison-Wesley.

Roessingh, J.J.M., Kappers, A.M.L & Koenderink, J.J. (2002). Forecasting the learning curve for the acquisition of complex skills from practice. NLR Technical Publication NLR-TP-2002-446. National Aerospace Laboratory, Amsterdam, the Netherlands.

Teague, R.C., Gittleman, S.S. & Park, O. (1994). A review of the literature on part-task and whole-task training and context dependency. Technical Report TR1010 (AD A285 954). US Army Research Institute for the Behavorial and Social Sciences, Alexandria, VA.

Vidulich, M.A., McCoy, A.L. & Crabtree, M.S. (1995). Attentional control and situational awareness in a complex air combat simulation. Paper presented at AGARD Symposium on “Situation Awareness”. Published in AGARD report CP-575. Neuilly-sur-Seine, France.

Wightman, D.C. & Lintern, G. (1985). Part task training for tracking and manual control. *Human Factors*, Vol. 27(3), 267-283.

THE IMPACT OF AUTOMATION ON TEAMWORK IN AIR TRAFFIC CONTROL

Jan Joris Roessingh
Rolf Zon
National Aerospace Laboratory NLR
Amsterdam, the Netherlands

Oliver Straeter
Eurocontrol
Brussels, Belgium

Ryoko Fukuda
Technische Universität
München, Germany

Generally spoken, teamwork may be at considerable risk when new automated technologies are introduced at the Working Positions of operators. Automation may replace team functions, team structure and changes the composition of the team. Team roles are unavoidably redefined and communication patterns are altered. For Air Traffic Control, EUROCONTROL developed measures for the impact of new computerised systems (automation) on teamwork. This was done within the framework of EUROCONTROL's SHAPE project.
In this paper we describe the development and experimental test of one of these measures (the "teamwork questionnaire") in a Tower Control environment. We propose a method for the validation of this questionnaire and focus on questionnaire-items that could be validated with Eye Point-of-Gaze data of the team members

Introduction

Generally spoken, team tasks are at considerable risk when automated technologies are introduced. Automation effects operating at the individual level may have a gamut of effects when distributed across teams.

As automation entirely or partially replaces team functions, team structure and changes the composition of the team, team roles are unavoidably redefined and communication patterns are altered (Bowers et al., 1993; Wiener, 1993; Jentsch et al., 1995; Bowers et al., 1996; Mosier & Skitka, 1996). While in the past it was assumed that workload would decrease with the introduction of automation, this advantage has been only partly realised. Automation substitutes human activities by 'machine activities' in combination with new human activities, while not leading to lowered workload levels. Additionally, Situation Awareness (SA) may decline as a result of (1) monitoring demands and subsequent vigilance decrements, (2) complacency due to over-reliance on automation, (3) system complexity, (4) poor interface design, (5) inadequate training or (6) lack of trust in automation (Endsley, 1997; Paris et al., 2000).

Goal & Scope

The 'Solutions for Human-Automation Partnerships in European ATM (SHAPE)' Project addresses the challenges on human factors as a consequence of the introduction of automation in ATM. These challenges concern:

- the level of trust that controllers have in automated tools;
- the effect on the controllers' situation awareness when using automated tools;
- the changes in skills needed to perform the controllers' job;
- the recovery from system failures when these occur in automated systems;
- the changes in (mental) workload that result from working with automation;
- the level of support needed when older controllers make the transition to a system with a higher level of automation than the one previously used;
- the changes in teamwork when a team of controllers make the transition to such a system.
–

The work presented in this paper is solely concerned with this latter point. To measure the changes in teamwork, a teamwork questionnaire was developed. This questionnaire is based on certain hypotheses about changes in teamwork, which will be clarified in the next section. Each hypothesis about teamwork underlies one or more questionnaire items. In addition a 'teamwork self-rating instrument', with which team members assess certain aspects of their own teamwork, and a 'teamwork observation instrument', with which external observers assess different aspects of teamwork, were developed. The latter two instruments will be used in an attempt to cross-validate some of the questionnaire items. In addition, eye-tracking measures, or more specifically, Eye-Point-Of-Gaze (EPOG) measures will be used, also to cross-validate some of the questionnaire items.

Teamwork in ATM: Definition

A team is defined as a group of two or more people who interact dynamically, interdependently, and adaptively toward a common and valued goal/objective/mission, who have each been assigned specific roles or functions to perform, and who have a limited life span membership (Salas, Dickinson, Converse & Tannenbaum, 1992).

In Air Traffic Control (ATC) this common goal is a safe and efficient control of traffic, in accordance with procedures and agreements. In principle, one could think of an ATC team, as the team consisting of the controllers of a specific sector and, in addition, the controllers of adjacent sectors, the cockpit crew of aircraft under control and possibly other personnel (such as ATC-system maintenance technicians) present in the control room. However, to keep matters simple and because this project concerns teamwork measures that are applicable to a wide range of controller-in-the-loop simulation studies, the team is defined as consisting of the controllers which are together responsible for one specific sector or aerodrome area, only. Hence, the focus of the to-be-developed teamwork concept is on the team on the ground. We define teamwork as the seamless integration of specific skills, knowledge and attitudes that allow team members to adapt and optimise their performance. We define a skill (or ability) as a goal-directed and well-organised behaviour that is acquired through practice. An example of a teamwork skill is a controller's ability to predict the behaviour of other team members in a developing traffic situation. Such a skill enables the team member to optimally support the team. In this paper we consider a teamwork skill as an individual skill, not the skill of the team as a whole. However, when assessing teamwork skills, it may be straightforward to analyse the behaviour of the team as a whole (for example, when analysing team communications), without being able to assess the skill at the individual level. This is not considered problematic, because the goal of the current project is to refine and validate instruments that measure the impact of automation on teamwork, rather than assessing individual skills. The basis for the acquisition and fine-tuning of teamwork skills is suitable knowledge and attitudes with respect to teamwork. Knowledge is difficult to define, but generally the following building blocks are recognised: (1) declarative knowledge (facts and concepts), (2) procedural knowledge: procedures and strategies, and (3) conditional knowledge: principles and conditions.

Examples of teamwork knowledge in each of these different building blocks are (1) understanding one's own function in the team, (2) knowledge of communication strategies such as ways to give and receive feedback and constructive criticism, and (3) the principles and conditions for creating and retaining a good teamwork atmosphere.

Teamwork attitudes are defined as an internal state that influences a team member's choices or decisions to act in a particular way (Cannon-Bowers *et al.*, 1995). Two examples of teamwork attitudes are (1) belief in the importance of teamwork and (2) belief in continuous learning as one of the main functions of the team.

Impact of Automation on Teamwork in ATC

Examples of team tasks in ATC are:

- Taking over the working position from another controller;
- Make others aware of, for example, unusual tracks of aircraft;
- Monitor fellow team members for performance, SA, and workload;
- Engage in (de-)briefings.
-

One way to express the impact of automation on team tasks is by defining to what extent the task has been taken over by a computer application.

Sheridan and Verplanck (1978) first proposed ten possible levels of allocation of decision-making tasks between humans and computers. More recently, Parasuraman, Sheridan and Wickens (2000) reconsidered a model of independent information processing functions and on that basis proposed a set of ten Levels Of Automation (LOAs):

1. the computer offers no assistance;
2. the computer offers a complete set of decision alternatives;
3. the computer narrows the selection down to a few;
4. the computer suggests an alternative;
5. the computer executes the suggestion if the human approves;
6. the computer allows the human a restricted time before automatic execution;
7. the computer executes automatically, then necessarily informs the human;
8. the computer informs the human only if asked;
9. the computer informs the human only if it (the

computer) decides to;

10. the computer decides everything and acts autonomously, ignoring the human;

In their report on the future of air traffic control a panel of the US National Research Council (NRC, 1998) recommended that, for system functions with relatively little uncertainty and risk, a high LOA is appropriate. However, when the system function is associated with greater uncertainty and risk, the LOA should not be more than level 4 (i.e. 'the computer suggests an alternative'). The panel adds to this recommendation: "Any consideration for automation at or above this level must be designed to prevent: loss of vigilance, loss of situation awareness, degradation of operational skills, and degradation of teamwork and communication. Such designs should also ensure the capabilities to overcome or counteract complacency, recover from failure, and provide a means of conflict resolution if loss of separation occurs." (NRC, 1998). In the SHAPE Project, a different classification of LOA is proposed. In contrast to the more general LOAs adopted by the NRC Panel, these LOAs are more specific for ATC systems. An application (or system component) is evaluated on *six dimensions* as follows:

1. The automation features that may facilitate *information extraction* are automatic highlighting, cueing, de-cluttering and filtering.
2. The automation features for *information integration* are automatic arranging and prioritisation.
3. The automation features for *information comprehension* are automatic comparison, diagnosis, prediction and testing.
4. The automation features for *decision/choice* are automatic option generation, option prioritisation, evaluation of options and option choice.
5. The automation features for 'response execution' (or 'action implementation') are automatic input (e.g. voice recognition), output (e.g. speech synthesis), implementing a response and implementing an emergency response (the system judges, according to known rules, that an action/response is required).
6. The automation features for 'information retention' are automatic reminders, history tracking and auto-delete.

When all of the above mentioned features are present, the LOA of the component is the highest possible.

Team Skills, Knowledge and Attitudes

On the basis of task analysis, the most import team skills, knowledge and attitudes were identified. The team skills were categorised in "co-operation skills" and "co-ordination skills", and further subcategorised in "leadership skills" and "followership skills".

<table>
<tr><th colspan="4">SKATE
Skills, Knowledge and Attitudes for TEamwork</th></tr>
<tr><th colspan="4">Core Skills</th></tr>
<tr><td colspan="2">Team co-operation
(team process oriented)
• Monitor/assess each other's performance;
• Monitor/assess each other's workload;
• Monitor/assess each other's SA;
• Predict each others behaviour;
• Prevent future overloading of the team;
• Back-up others to prevent overloading;
• Adjust to each other's working style.</td><td colspan="2">Team co-ordination
(outcome oriented)
• Ensure which tasks are:
✓ Entirely own responsibility;
✓ Shared with other team members;
✓ Performed by the automated system.
• Assess the traffic situation / seek information;
• Provide updates on the traffic situation;
• Co-ordinate with other team members;
• Make team decisions;
• Select course of actions;
• Synchronise team actions.</td></tr>
<tr><th>Leader type qualities</th><th>Follower type qualities</th><th>Leader type qualities</th><th>Follower type qualities</th></tr>
<tr><td>• Redistribute workload when required;
• Take initiative;
• Motivate the team;
• Tactful alert other members to mistakes;
• Suggest ways to find and remedy errors.</td><td>• Exhibit assertiveness;
• Ask for assistance when necessary;
• Help others with difficult tasks or problems.</td><td>• Focus the team on its tasks;
• Form and disseminate plans;
• Assign tasks;
• Give orders and directives;
• Take command.</td><td>• Accept suggestions or criticisms;
• Perform self corrections;
• Provide an input or response when asked.</td></tr>
<tr><th colspan="4">Attitudes</th></tr>
<tr><td colspan="4">General attitudes towards teamwork:
• Belief in the importance of team work.
• Belief in the continuous learning process as one of the main functions of the team.</td></tr>
<tr><th colspan="2">Co-operative attitudes</th><th colspan="2">Co-ordinated attitudes</th></tr>
<tr><td colspan="2">• Team spirit, team morale, team cohesion;
• Willingness to maintain identity of the team.</td><td colspan="2">• Shared vision;
• Mutual trust in the team.</td></tr>
</table>

Figure 1. *The SKATE model*

The resulting categorisation, i.e. the SKATE model (Skills, Knowledge and Attitudes for Teamwork) is depicted in Figure 1.

Measures. All components of the SKATE Model, which represent aspects of teamwork that can possibly be affected by automation, are covered by three measures (paper forms) that were developed in the current project. These are:

- Observation form - O: A form that allows (external) observers to rate a number of observable teamwork aspects that take place within the team when the team is performing their (automated) task. This observation form is importantly based on Entin & Entin (2001).
- Questionnaire - Q: A form that the team members will complete after participation in experimental trials with the new automation. Each team members provide information about the impact of the automation on his/her teamwork skills, knowledge and attitudes.
- Self-rating form – R: A form that will be completed by team members after each experiment run. The team members provide information about team workload and team

situation awareness during the (simulator) experiment. In addition, the team members give ratings for the fidelity of the experiment run (e.g. whether the traffic sample is realistic) which allows the experimenters to do a more detailed analysis.

Experimental Validation

Method

General. The experiment took place in NLR's high fidelity Tower Research Simulator (TRS, see Figure 2). Details about the simulator itself can be found in Zon and Roessingh (2004). The team of controllers went through a series of experimental runs. In half of the runs, an automation application (i.e. Collaborative Decision-Making or CDM software) was switched ON, in the other half of the runs the application was switched OFF.

Collaborative Decision-making (CDM) refers to a set of applications aimed at improving flight operations through the increased involvement of (1) airspace users, (2) ATM service providers, (3) airport operators and (4) other stakeholders in the process of air traffic management. Collaborative decision-making applies to all layers of decisions, from longer-term planning activities to real-time operations, and is based on the sharing of information about events, preferences and constraints.

In this case, when the CDM is switched ON (i.e. the ON condition) a number of scheduling tasks, are dealt with by the system. In the OFF condition there are more planning tasks that the controllers (particularly the Departure Planner) have to perform themselves.

Subjects and task. Three different air traffic controllers participated in the experiment. Each of them had a different task while together they formed a team that guided aircraft from the gate to a runway and vice versa. Aircraft waiting at the gate were handed-off by the Departure Planner. Subsequently, the aircraft were guided by the ground controller and finally sent to the runway for take-off by the tower controller. Arriving aircraft were, via the tower controller, passed onto the ground controller and then at the gate were waiting for their next departure as planned by the departure planner. So, in fact, teamwork in the tower is rather serial in nature, with the hand-over of aircraft from one controller to the next.

The Departure Controller (DC) performs his tasks as follows. The DC is facing a display (see Figure 3), which, at the top, displays the Electronic Flight Strips (EFSs) of the aircraft that will soon come under his control. In the middle, the EFSs of aircraft that are currently under his control are displayed. At the bottom it shows the EFSs of aircraft that the DC has handed over to the ground controller. Apart from the usual flight strip information, also advanced time planning and scheduling information is displayed. In the ON condition, the CDM software schedules the departing aircraft automatically.

Figure 2. *Controllers are seated at their working positions in NLR's Tower Research Simulator. From Left to Right: Departure Planner, Tower Controller and Ground Controller.*

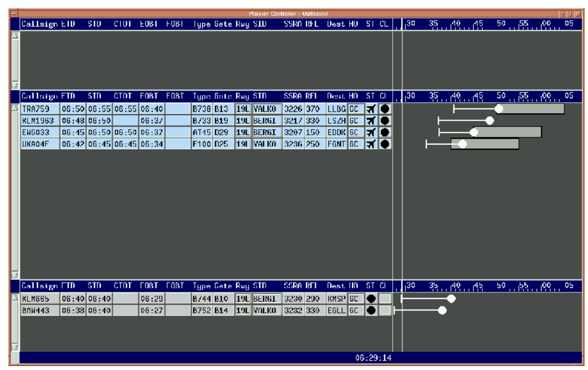


Figure 3. *Display of the Departure Planner, showing Electronic Flight Strips (EFSs).*

Based upon the status that on the display, the DC instructs pilots and allows them to start up. The fewer time the DC needs to spent looking at this monitor with electronic flight strips, the more time he can spent either looking outside or at other displays. By looking outside, the DP can check whether the aircraft indeed adhere to the planning. Other displays include the arrival departure table (which provides the DC with the possibility to look ahead in time) or the map display depicting ground traffic at the airport.

Procedure. The controllers had to control traffic samples during experimental runs of approximately three hours each. Each experimental run took place in a different condition (OFF and ON). In the ON condition, CDM tools helped the controllers with scheduling, in order to adjust their work to the task-load of their colleagues.

"Use of runways" was an additional variable. Either one runway was in use as a mixed runway (one runway in use for both departures and arrivals) or two runways were in use as segregated runways (the first one dedicated to arrivals and the second one dedicated to departures). A final additional variable was the traffic load in the different traffic samples. The traffic samples were designed to regulate task-load. In each experimental run, a low to medium taskload was applied since the aim was to give the ATCos the opportunities to test the system and not to 'force' them into handling as many aircraft as possible.

During all experimental runs each controller had the same role in the team. An overview of the different experiment runs is given in the Table below.

	Scenario no.	CDM OFF	CDM ON
Segregated	1S	Run 1	Run 5
	2S	Run 6	Run 4
Mixed	2M	Run 3	Run 2

Questionnaire Q

After all experimental runs, the questionnaire Q was administered. Questionnaire Q consisted of 33 question items. All of the items were put in the form of statements regarding the new automation application (CDM) that the participants had experienced in half of the experiment runs. The participants were asked to rate on a 5-point scale to what extent they agree with each statement.

Eye-Tracking Equipment

For validation of a number of items from questionnaire Q, eye-tracking equipment was used. The eye-tracking behaviour of the Departure Controller was measured with a so-called GazeTracker (Mooij & Associates, 1996). This equipment measures the track of the Eye-Point-Of-Gaze (EPOG) on predefined Areas of Interest (in this case, the Areas of Interest are all objects the DP could possibly look at, including the three computer displays mentioned earlier, the outside world, his fellow team members and his desk surface).

The duration that the DP gazes at a particular Area of Interest is called the 'dwell time'. In addition to the dwell times, the scanning pattern, amounts of fixations, pupil diameter and eye-blink activity (that permits blink rate, duration and other measures to be derived) of the departure controller's left eyes were recorded as indicators of mental and visual workload (see Harris et al., 1986; Stern et al., 1984; Stern & Kelly, 1984; Stern, 1994; Goldstein et al., 1985; Wilson et al., 1987, 1993). It is assumed that there is a negative correlation between (visual) workload and eye blink rate. The scanning behaviour is considered to be an indicator of the DC's mental state and focus of attention. It was generally assumed that when the DC was looking at a particular Area of Interest, that he was paying attention to it (or an object in it).

During the experiment runs, the scene in the tower was videotaped and separate sound-recordings were made of the voice communication between controllers (intra-team communication) and the communication of the DP with pilots of aircraft under his control.

Validation Methodology

For each questionnaire item, hypotheses relating the rating of the controllers and measurable behaviour were formulated. In other words, hypotheses regarding the expected effect of the team performance in the measured variables were formulated for each item of the questionnaire. Separate hypotheses were formulated for those situations where automation was ON or OFF.

On this basis, hypotheses underlying questionnaire items could be validated on a three-point scale:

- Validated: The key measures fully support the rating of an item.
- Not confirmed: The key measures do not contradict the rating of an item. However, further study is needed to validate the item. It is well possible that the hypothesis concerning this item is true, and that there are reasons to assume that the item can be validated in the future.
- Contradictory: The key measures do contradict the rating of an item.

Results

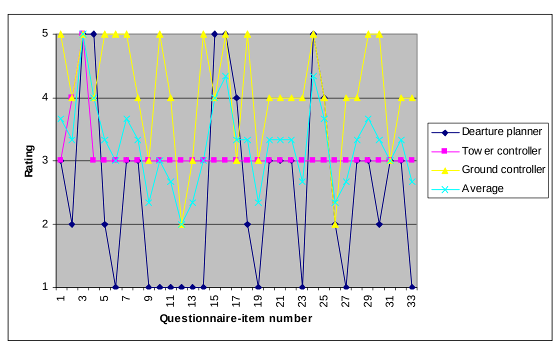


Figure 4. *Ratings of all team members on 33 questionnaire items.*

The ratings that the controllers gave on the to be validated questionnaire Q are visualised in figure 4. Notice the differences between the three different types of controllers.

An overview of the recorded EPOG data is given in Table 1 and 2 . The entire set of questionnaire items, associated measurements and validations can be found in Zon and Roessingh (2004). However an example of an interesting insight that was gained based upon eye tracking data is the fact that the DP made more fixations on the arrivals and departures tables under the CDM OFF conditions, when compared to the CDM ON conditions. It is apparent that in the ON condition, the CDM application takes over some planning and scheduling tasks from the DP, such that the DP can give more attention to other Areas of Interest (such as the outside world). This is the type of result that could be derived from eye-tracking data summarised in Tables 1 and 2.

Table 1. *EPOG for two complete (comparable) runs (segregated runway use)*

Run 4 **(automation on, scenario 2, segregated runway use)**	**Run 6** **(automation off, scenario 2, segregated runway use)**
Blink rate (blink/min) 48.41	Blink rate (blink/min) 24.5
Fixations on arrivals table 2.15 %	Fixations on arrivals table 7.16 %
Fixations on departure table 15.15 %	Fixations on departure table 15.24 %
Fixations on EFS top 3.56 %	Fixations on EFS top 0.98 %
Fixations on EFS middle 2.61 %	Fixations on EFS middle 27.59 %
Fixations on EFS bottom 0.05 %	Fixations on EFS bottom 0.61 %
Fixations on EFS right 4.49 %	Fixations on EFS right 8.91 %
Fixations on airport map 5.5 %	Fixations on airport map 8.48 %
Fixations on outside (projection screen) 17.04 %	Fixations on outside (projection screen) 10.01 %
Fixations on other predefined areas 9.31 %	Fixations on other predefined areas 2.62 %

(EFS stands for: Electronic Flight Strip)

Table 2. *EPOG for two complete (comparable) runs (mixed runway use)*

Run 2 **(automation on, scenario 2, mixed runway use)**	**Run 3** **(automation off, scenario 2, mixed runway use)**
Blink rate (blink/min) 41.10	Blink rate (blink/min) 8.64
Fixations on arrival-departure table 10.49 %	Fixations on arrivals table 22.04 %
Fixations on EFS top 1.91 %	Fixations on EFS top 0.29 %
Fixations on EFS middle 1.64 %	Fixations on EFS middle 8.63 %
Fixations on EFS bottom 0.02 %	Fixations on EFS bottom 0.61 %
Fixations on EFS right 2.86 %	Fixations on EFS right 5.66 %
Fixations on airport map 1.66 %	Fixations on airport map 1.75 %
Fixations on outside (projection screen) 14.07 %	Fixations on outside (projection screen) 1.3 %
Fixations on other predefined areas 10.96 %	Fixations on other predefined areas 8.75 %

(EFS stands for: Electronic Flight Strip)

A number of questionnaire items of questionnaire could be validated. There were also items that, for statistical reasons, could not be validated.

Discussion and Conclusion

General

The aim of the work was to develop a measurement battery for human factors aspects of automation. In particular it validated a set of teamwork measures. First the model for Skills, Knowledge and Attitudes in Teamwork (the SKATE model) was developed. The components of the SKATE model formed the base for the development of a set of paper and pencil instruments. One of these instruments – the questionnaire Q - was validated in a realistic simulator experiment. The other instruments and eye-tracking data were used as references against which questionnaire Q was validated.

Teamwork in automated systems is more than just voice communication. The study on a Collaborative Decision Making system in the Tower Simulator showed clearly that usual communication channels, usually suitable for recovering from critical situations, are lost. As such, the Observation form (O), which was developed to structurally observe spoken intra-team communication was of little use, because with CDM, team members didn't share their intentions, plans and decisions by speech anymore, but just via the Human Machine Interface. In addition, the experimental setting probably interfered with natural communication between the ATCos, which probably resulted in considerable less communication than in real-life. Automated systems cannot substitute speech act and other means are needed to overcome the loss of non-verbal communication (i.e. the richness of face-to-face communication in particular) regarding, emotional states, workload and cognitions (e.g. when diagnosing system disturbances or unusual tracks).

Designers of systems therefore have to carefully take into account teamwork. This part of system design needs as much attention as the behaviour of the individual operator. The currently developed set of teamwork instruments is one means that can be used to assure that system design of ATC applications meets teamwork needs of controllers.

Eventually seventeen out of 33 items could either be verified or validated in the described experimental runs and with the adopted validation approach. Hence, a second validation step is required before the questionnaire for full validation of the questionnaire.
Finally, two remarks need to be made when interpreting the outcome of the current study. First, information exchange is increasingly becoming a system task, rather than solely the domain of team members. This development should be taken into account in future teamwork measurement instruments. Therefore, there was insufficient data to validate the Observation form (O). Second, validation in the statistical sense, e.g. rejection of general hypotheses about the impact of automation on teamwork with a certain percentage of confidence, was impossible. All results are based upon trends.

References

Bowers, C.A., Deaton, J., Oser, R., Prince, C. & Kolb, M. (1993). The impact of automation on crew communication and performance. In: R. Jensen (Eds), *Proceedings of the Seventh International Symposium on Aviation Psychology*, 758-761, Columbus, OH, US: Ohio State University.

Bowers, C.A., Oser, R.L, Salas, E. & Cannon-Bowers, J.A. (1996). Team performance in automated systems. In: R. Parasuraman & M. Mouloua (Eds), *Automation and Human Performance: Theory and Applications*, 243-263, Mahwah, US: Lawrence Erlbaum.

Cannon-Bowers, J.A., Tannenbaum, S.I., Salas, E. & Volpe, C.E. (1995). Defining team competencies and establishing team training requirements. In: R. Guzzo & E. Salas (Eds), *Team effectiveness and decision-making in organizations*, 333-380. San Francisco, CA, US: Jossey Bass.

Entin, E.E. & Entin, E.B. (2001). Measures for evaluation of team processes and performance in experiments and exercises. In: *Proceedings of the 6th International Command and Control Research Technology Symposium*, June 19-21, Annapolis, MD, US.

Goldstein, R., Walrath, L.C., Stern, J.S. & Strock, B.D. (1985). Blink activity in a discrimination task as a function of stimulus modality and schedule of presentation. *Psychophysiology*, 22, 629 - 635.

Harris, R.L., Glover, B.J. & Spady, A.A. (1986). Analytical Techniques of Pilot Scanning Behavior and Their Application. *NASA Hampton Technical Paper 2525*.

Jentsch, F., Bowers, C., Redshaw, B., Bergen, H., Henning, J. & Holmes, B. (1995). Differential effects of automation on team performance, workload and communications. Technical Report, Naval Air Warfare Centre - Training Systems Division, FL, US.

Mosier, K.L. & Skitka, L.J. (1996). Human decision-making and automated decision aids: made for each other? In: R. Parasuraman & M. Mouloua (Eds), *Automation and Human Performance: Theory and Applications*, 201-

220. Mahwah, US: Lawrence Erlbaum.

NRC (1998). *The future of air traffic control - Human operators and automation*, by C.D. Wickens, A.S. Mavor, R. Parasuraman & J.P. Magee (Eds). National Research Council (NRC), Washington DC, US: National Academy Press.

Parasuraman, R., Sheridan, T.B. & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, *Man and Cybernetics*. Part A: Systems and Humans, 30(3), 286-297.

Paris, C.R., Salas, E. & Cannon-Bowers, J.A. (2000). Teamwork in multi-person systems: a review and analysis. *Ergonomics*, 43-8, 1052-1075.

Salas, E., Dickinson, T.L., Converse, S. & Tannenbaum, S.I. (1992). Towards an understanding of team performance and training. In: R. Swezey & E. Salas (Eds), *Teams, their training and performance*, 3-29. Ablex, Norwood.

Sheridan, T.B. & Verplank, W.L. (1978). Human and computer control of undersea teleoperators. *Technical report*. MIT Man-machine systems laboratory, Cambridge, MA, US.

Stern, J.A. (1994). Blink Rate: A Possible Measure of Fatigue. *Human factors*. 36(2), 285-297.

Stern, J.A. & Skelly, J.J. (1984). The eyeblink and workload considerations. *Proceedings of the Human Factors Society 28th Annual Meeting*, San Antonio, TX, US.

Stern, J.A., Walrath, L.C. & Goldstein, R. (1984). The Endogenous Eyeblink. *Psychophysiology*. 21(1), 22-33.

Wiener, E.L. (1993). Crew Co-ordination and training in the advanced technology cockpit. In: E.L. Wiener, B.G. Kanki & R.L. Helmreich (Eds), *Cockpit Resource Management*, 199-229. San Diego, CA, US: Academic Press.

Wilson, G.F., Badeau, A. & Gawron, V. (1993). Evaluation of two inflight sensors using heart-rate and eye-blink measures. *Aviation Psychology*.

Wilson, G.F., Purfis, B., Skelly, J., Fullenkamp, P. & Davis, I. (1987). Physiological data used to measure pilot workload in actual flight and simulator conditions. In: *Proceedings of Human Factors society - 31st annual meeting*.

Zon, G.D.R. & Mooij, H.A. (1996). GazeTracker for Real-time, Omni-directional and Digital Point-of-gaze Determination - Installation and Operation Guide - EPOG Version 4.3, MA-95-15, December.

Zon, G.D.R. & Roessingh J.J.M. (2004). A Measure to Assess the Impact of Automation on Teamwork. Contract Report NLR-CR-2004-173, NLR, Amsterdam, the Netherlands.

NEAR-TO-EYE DISPLAY AND AUGMENTED REALITY CONCEPTS FOR AIR TRAFFIC TOWER CONTROLLERS: ISSUES AND CHALLENGES

Dr. John W. Ruffner, **Dr. Jim E. Fulbrook**
DCS Corporation
Alexandria, Virginia

Marc Foglia
NVIS, Inc.
Reston, Virginia

Tower controllers are responsible for maintaining safe separation between airborne aircraft in the airport traffic control area, and separation between aircraft, equipment, and personnel on the airport surface. In this paper, we summarize recent work to develop and evaluate user-acceptable hardware and software solutions that will reduce diversions and augment or enhance controller capabilities, especially in limited visibility conditions. We characterized controller tasks where a near-to-eye display and augmented reality techniques can aid controller performance, and identified form factor variables that influence user acceptability of hardware configurations. We developed an out-the-window concept of operation and analyzed the hardware requirements and feasibility of three near-to-eye viewing systems: two head-mounted monocular displays and a held-to-head binocular display. When fully developed, these display systems should enhance tower controller situation awareness, and reduce such distractions as having to frequently attend to and respond to head-down (console) display information. There are potential users of this display system concept in all branches of the military services and in the commercial sector, and potential utility for surface surveillance operations in support of homeland security, law enforcement, search and rescue, firefighting, and special operations.

Introduction

Air traffic controllers in the tower environment are responsible for control of traffic on the ground and in the air within the airport traffic control area. Ground control, departure and arrival sequencing, and surface management are continuous challenges. Tower controllers live in an information-rich world, processing data from a multitude of sources. The controller must maintain situation awareness (SA) while assimilating information from such sources as out-the-window observations; scanning of head-down displays; audio management and interactions between aircrews; departure, arrival, and traffic pattern management; ground operations management, interaction between other controllers, and flight data strip management.

Controllers must frequently divert their attention away from the external scene, which could impair safe and effective operations, as the controller is the ultimate decision and management authority on an airfield. In this paper we discuss recent work we conducted in this area that was sponsored by the Air Force, and which built on the results of our previous work with the Federal Aviation Administration (FAA). The focus of the effory is to develop and evaluate user-acceptable hardware and software solutions, using near-to-eye display and augmented reality (AR) techniques and concepts, that will help reduce diversions and augment or enhance tower controller capabilities, especially at night and in limited visibility conditions.

Tower Controller Positions

The two key tower controller positions in military and civilian towers are the *Local Controller* (LC) and the *Ground Controller* (GC). The LC is primarily responsible for handling arriving and departing traffic at the airport. The LC area of responsibility includes the active runways and the airspace within a 5-mile wide radius of the airport. Generally, the LC position interfaces with Ground Control and related tower positions. The GC is primarily responsible for directing aircraft to and from the runway. The GC is also responsible for directing other aircraft/vehicular movement on the airport movement area and disseminating information to support operations (e.g. traffic, weather, equipment status, delays/flow, flight plans, etc.). Clearly, the tower environment is a dynamic environment in which the LC and GC must adjust traffic flows, evaluate new information, and closely coordinate and communicate their efforts.

Information Sources

Tower controllers obtain information required to perform their duties from many sources. Information on individual flights and their intended airborne or ground path is transferred to the controller through the use of flight progress strips, surface map displays, and verbal communications. The primary source of information regarding aircraft location is the out-the-window view from the tower. However, the controller must continually monitor a large number of console displays that provide information on local weather conditions, as well as arrival and departure information that must be relayed to pilots as needed. This must be correlated with communications between the controller and the aircraft or other air traffic control (ATC) facilities.

To acquire and maintain situation awareness, controllers must know at a minimum, aircraft and ground vehicle identity, location, and intent (Piccione, Krebs, Warren, and Driggers, 2002). A critical part of the task of knowing aircraft location and intent is the ability to identify the specific aircraft that are being controlled. This capability is augmented at some facilities with radar displays that show the identity and location of aircraft in the airspace surrounding the airport. At larger airports, ground surveillance radar provides information regarding objects on the airport movement area. An example of a ground surveillance radar display is the current Airport Surface Detection Equipment-X Series system (ASDE-X).

Flight progress strips are used at some facilities as an analogy for the flight and are manipulated on the console during the hand-off process between tower positions. They provide detailed flight information for each departure aircraft, including the aircraft type, first departure fix, flight plan, and flight identification (ID) of the aircraft. The strips are marked with updated information as an additional means of information storage and transfer. Taken together, the ASDE-X map display, flight strips, and the Digital Bright Radar Indicator Tower Equipment (D-BRITE) display (a repeater display of the terminal radar control [TRACON] display) provide a good picture of the current state of the terminal airspace and airport surface, and help the tower controller build situation awareness from multiple look-down sources.

The controller must use a scan pattern outside and inside the tower to assimilate, correlate, and integrate information to build and maintain situation awareness. The information inside the tower is presented on a variety of displays that may be imbedded in the console, placed on the console as a freestanding unit, or mounted overhead of the tower windows. The controller must determine what information is needed, retrieve that information from displays throughout the tower cab, and mentally integrate the information. Tower controllers frequently cite problems associated with the requirement to use large scan patterns inside the cab that detract from their out-the-window task of monitoring the airspace, runway, and airport movement area (taxiways and ramps). These types of typical display options currently used in a control tower all require distance viewing across a wide field of regard. This presents a challenge where user-centered solutions that reduce look-down time and improve information management for controllers would be beneficial.

Effects of Reduced Visibility

The out-the-window scene is severely degraded during night and limited visibility conditions. When visibility is restricted, controllers may be able to maintain some degree of SA by following established procedures and forming expectations of key events. However, their overall SA is still significantly degraded. Controllers must establish and maintain a mental image of the airport layout, and use graphical aids (e.g., taxiway diagrams) and position reports to determine the location of aircraft and other objects on the surface, and form expectancies of where an aircraft or vehicle should be (Piccione et al., 2002).

Potential Solutions

Two potential solutions for increasing SA, enhancing safety, and increasing throughput under daytime and limited visibility conditions are (1) to supplement the controller's visual capabilities with an Enhanced Vision System (EVS), and (2) to provide the controller with a display with text and symbology overlaid on key elements of the out-the-window, video, or EVS scene to augment the perception and understanding of the scene. The visible or EVS scene can be presented on such display devices as a head-mounted monocular display (HMMD) or a held-to-head binocular display (HHBD). An EVS can restore some of the critical visual capability that may be lost or severely reduced due to darkness or reduced visibility.

A recent study by the FAA (Piccione et al., 2002) investigated the use of electro-optic sensors to enhance tower controller visual capabilities during poor atmospheric or low-illumination conditions. The field data and modeling results suggest that using a long wave infrared (LWIR) sensor could improve controller nighttime detection, recognition, and identification of obstacles/targets on the airfield surface. Critical issues included the sensor's field-of-view, the field of regard, the mechanism for mounting the sensor(s), the display medium (e.g., head-down vs. head-mounted), and the need for a head-tracking system versus fixed sensor (camera) positions.

Need for a New Display Paradigm

Tower controllers may benefit from a near-to-eye augmented reality display that allows a continuous head-up, out-the-window view of the runway and eliminates, or greatly reduces, the time-consuming scanning, frequent eye accommodation changes, and cognitive integration currently required to access this same data on head-down displays. We examined what are called "near-to-eye" displays because such systems provide electronic and miniaturized viewing capabilities in a display placed generally within one inch from the viewer's eye.

The use of a near-to-eye held-to-head or head-mounted augmented reality display would allow the presentation of context-sensitive information and the "scene-linking" of text or imagery that can cue the presence of aircraft and highlight the location of runways, thus improving overall ground safety. As discussed previously, these safety benefits can be significant in low visibility conditions, in which scene-linked imagery may highlight the location of planes or vehicles on a visible video or EVS mage that the controller may not be able to otherwise see directly.

Research Objectives

Our research has three objectives:

(1) To understand tower controller surveillance tasks, and how near-to-eye displays and AR and EVS techniques relate to these tasks.
(2) To analyze the technical requirements (e.g., tracking, resolution) so a feasible design of required display capability can be developed.
(3) To understand the form factors as well as the technical and social challenges for implementing such a display system.

Technical Approach and Findings

Tower Controller Tasks

Our first task was to gain an understanding of the tasks performed by tower ground controllers and local controllers, the information they need to perform these tasks, and the sources of this information. We reviewed the available technical literature to identify relevant studies. Key findings from earlier studies, as well as findings from our previous work on EVS requirements for tower controllers, are discussed in Ruffner, Deaver, and Henry (2003) and Ruffner, Fulbrook, and Foglia (2004). Relevant findings from two recent studies sponsored by the FAA and the National Aeronautics and Space Administration (NASA) are summarized below.

*FAA Tower Controller Study***.** The FAA recently conducted a study to examine factors contributing to the complexity of tower controller tasks (Koros, Della Rocco, Panjwani, Ingurgio, and D'Arcy, 2003). This study produced data on tower controller decision-making strategies, information requirements, and information sources for both the LC and GC positions.

The most important information *elements* common to the LC and GC, in descending order of importance, were: (1) aircraft position, (2) aircraft identification, and (3) route to be followed during taxi operations. The most common information *sources* were: (1) out-the-window visual observation, (2) flight strips, (3) communication with the pilot, and (4) the D-BRITE radar display. Visual observation was considered the first or second most important source of information for over 60% of the information elements.

*NASA Surface Management System Study***.** Under NASA sponsorship, researchers from Booze, Allen, and Hamilton, and Ohio State University conducted a human factors assessment of the developmental Surface Management System (SMS) (Hitt, Duley, Kressen, Mafera, Smith, and Spenser, 2002). SMS is being developed as a decision support tool that provides controllers and airline personnel with aircraft-specific information and predicted departure demand information.

Both LCs and GCs desired aircraft identification and flight-specific information to be presented via data blocks on a surface map display similar to the ASDE-X display. The specific information provided, as well as the desired area to be covered by the map display, depended on the controller position. This result reflects the controllers' need to have integrated information in one location. Display clutter (i.e., excessive text and graphical information) was identified as a critical issue Because tower controllers interact with each other frequently to exchange information, the respondents judged that the displays need to be clearly visible to all tower controllers, and that the display designs need to be standardized in their use of color-coding and symbology.

In addition, the SMS study identified the primary tasks and subtasks for the ground and local controller positions, and developed procedural flow diagrams for the LC and GC tasks. The project report describes the tasks to be executed and the times where decisions were required for the tasks (e.g., Maintaining Runway

Balance) and subtasks (e.g., Determine Delay to Runway Threshold). In addition, the report lists the information requirements (e.g., aircraft type, aircraft identification) for each task and subtask, and the source(s) from which information can be obtained (e.g., out-the-window, map display).

Technical Issues

There are several technical issues that must be resolved for a near-to-eye augmented reality tower controller display concept, capable of interfacing with an EVS, to prove feasible and practical. These issues include the minimum acceptable field-of-view (FOV), resolution, sensing and head tracking requirements, and the implementation strategy for selecting and superimposing text and symbology on the out-the-window display field-of-view.

Display Field of View and Resolution. Field of view and resolution are key parameters of any head-mounted display system, and often are traded off during the design decision making process. Studies of pilot performance with head-mounted displays (HMDs) indicate that wider FOVs generally result in better performance and situation awareness. Our analysis indicated that, in general the field of view for a tower controller HMMD/HHBD should be variable; a typical wide angle to telephoto range (e.g., 28 mm – 200 mm, or with a visual angle of approximately 10 to 100 degrees) is desirable. Both the HMMD systems and HHBD systems we evaluated appear to have sufficient FOV and resolution. Specifically the HHBD (NVIS Virtual Binocular™) has a FOV of 40 degrees diagonal, and the HMMDs have FOVs of 23 x 17 degrees (Microvision Nomad™) and 16 x 12 degrees (MicroOptical SV-6 PC Viewer™) respectively. All displays have a minimum of 800 x 600 pixel resolution. In short, all the displays appeared to have sufficient, effective visual presentations.

Augmented Reality. AR techniques allow the visualization of complex data by superimposing supplementary information relevant to the task at hand, which is referenced to the real world. AR display enhancements to support operator tasks include presenting cueing information to guide attention throughout the visual scene, and providing supporting textual or graphical information. AR displays let users see the surrounding real world and augment their view by overlaying 2-D or 3-D virtual objects on or near their real world counterparts to create the impression that virtual and real objects coexist (Azuma, 2001).

AR display issues include: (1) registration (aligning objects in the real and virtual scene), (2) sensing (detecting and identifying objects in the environment), (3) latency (lag between the display presentation of the actual and displayed event), and (4) head tracking. In a static AR environment, the real-world objects must be carefully modeled to capture their geometry so that virtual objects are properly aligned with real objects in the scene. In a dynamic AR environment, such as in the ATC tower, position and orientation of moving objects must be continually updated in the scene-graph so that virtual objects are correctly rendered and registered. The timeliness and accuracy of the information is of paramount importance (Martinsen, Havig, Post, Reis, and Simpson, 2003).

Display Symbology. An AR display concept involves superimposing text and screen-referenced or scene-linked symbology on an out-the-window scene similar to what is done with an aviation or automotive head-up display (HUD). A HUD eliminates, or at least minimizes, the need for refocusing and for extensive eye scan movements between panel-mounted instruments and the out-the-window visual scene. Dividing attention across stimuli belonging to separate "domains" or perceptual groups (e.g., a digital altimeter vs. a wire-frame outline of a tank linked to a feature in the visual scene) can lead to attention narrowing, This effect is reduced somewhat with scene-linked, or conformal symbology (Yeh and Wickens, 2001).

There are two key challenges to display/real-time imagery integration. The first is providing screen-referenced text and symbology that presents information related to the scenes and real-time events as they unfold during normal duty performance (e.g. wind direction). The second is the more difficult challenge of providing scene-linked text and symbology information that directly links an object or event appearing in the display with the text/symbology information as it is dynamically presented. An example is that if an aircraft is taxiing on the airfield and the display is directed to gaze on the aircraft, identifying symbology will be automatically presented and tagged to the aircraft. Moving the gaze to another aircraft will cause the display to recognize, retrieve, and present a new set of data.

Critical essential information for scene-linking includes: aircraft identification (ID), surface vehicle type/ID, aircraft position (runway, taxiway location), and flight plan data (departure runway, location fixes, destination, etc.). Achieving augmented reality capabilities in a near-to-eye display involves the integration of sources into an interactive and dynamic presentation that enhances a user's situation awareness and task capabilities without overloading the person. This scene-linking represents a level of technological capability that has not been reliably demonstrated for similar situations to date.

Development and Implementation Issues

Display Concepts. We investigated two different display concepts: (1) an optical see-through head-mounted display (see Figure 1) and (2) a held-to-head simulated binocular video see-through display (see Figure 2). Both display concepts can provide users with either screen-referenced or scene-linked symbology using near-to-eye display technology. The main difference between these two approaches is how the user interfaces with and interacts with the display. In the first case, the symbology is optically superimposed on the real world scene. In the second case, the symbology is superimposed on a video image of the real world scene. There are strengths and weaknesses for each approach, and technical challenges that must be overcome to make either one work (see Rolland and Fuchs, 2001).

Figure 1. *HMMD optical display concept.*

Form Factors and User Acceptability. We understand "form factor" here to mean the physical platform or mechanism that serves as the host for the display, or into which the display is attached or integrated. Even the most technologically sophisticated ATC display concept will not be used by tower controllers if it is too heavy, cumbersome, intrusive, or otherwise difficult to use. Accordingly, a fair question for either display concept is "Will controllers actually use one of these devices for extended periods?" The answer will likely be reduced to the issue of whether the advanced capabilities and benefits afforded by the technology offset the problems and costs induced by the encumbrance and potential sensory conflicts. Good human factors and ergonomic deign will be critical for achieving user acceptance.

Figure 2. *HHBD concept with thermal imagery.*

Augmented Reality Symbology Issues. A key issue is the type, amount, and placement of overlaid text and symbology, and the potential for information overload. There are guidelines for selecting and displaying imagery on aircraft HUDs for aircraft in-flight and surface operations (e.g., Mejdal, McCauley, and Beringer. 2001). However, it is not known how well the guidelines generalize to the tower cab environment. A new guideline development effort will likely be needed. Another issue is controller reaction to potential degradation or complete failure of a see-through HMMD or HHBD visual scene during operations.

Concept of Operations (CONOPS). Figure 3 and Figure 4 illustrate how the displays and symbology might appear and be used in an operational tower environment. Illustrated here are a daytime out-the-window situation (Figure 3) and night/low visibility condition situation using an EVS (Figure 4).

Conclusions

A near-to-eye display solution is feasible for air traffic tower controllers, especially when coupled with AR and EVS technologies. There are unique benefits to both HMMD and HHBD solutions and potential for the two systems to work together, as well as individually, in the tower cab environment. However, there are significant design, engineering, integration, and usability issues and challenges to achieving a solution that must be met. Our future efforts will involve developing a fully functional prototype AR display system, integrating the display system with available information sources, and conducting a usability assessment.

Figure 3. *Daytime display CONOPS.*

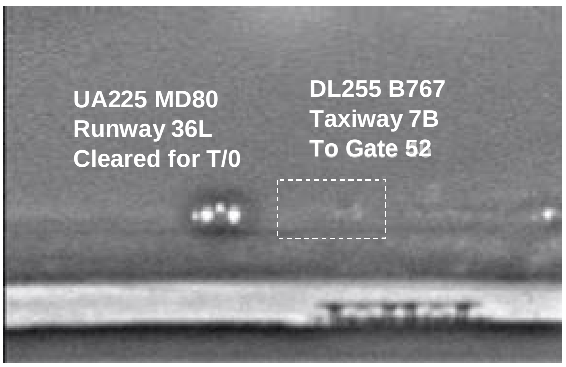


Figure 4. *Night/limited visibility display CONOPS*

Acknowledgments

This work was funded through an Air Force Phase I SBIR Phase I project, Contract Number F33615-03-M-6363, Dr. Peter Marasco, Air Force Research Laboratory, Wright Patterson AFB, OH (Technical Monitor), as a follow-on to FAA Contract GS-23F-0303K, Dino Piccione, (Technical Monitor). The authors acknowledge the assistance of Dr. D. Michael McAnulty, FAA, and Dr. James Hitt of Booz, Allen, and Hamilton. The statements made in this paper reflect the opinions of the authors, and do not represent the position of the U. S. Air Force, the FAA, or DCS Corporation.

References

Azuma, R. T. (2001). Augmented reality: Approaches and technical challenges," In W. Barfield, & T. Caudell, (Eds.). *Fundamentals of Wearable Computers and Augmented Reality*. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.

Hitt, J. M., Duley, J., Kressen, S., Mafera, P. Smith, P., & Spenser, A. (2002). *Air traffic management and system development and integration: Surface Management System human factors assessment,* National Air and Space Administration, Washington, DC.

Koros, A., Della Rocco, P. S., Panjwani, G., Ingurgio, V, & D'Arcy, J. F. (2003). *Complexity in air traffic control towers: A field study. Part 1 - Complexity factors.* (DOT/FAA/CT-TN03/14), Federal Aviation Administration William J. Hughes Technical Center, Atlantic City, NJ.

Martinsen, G. L., Havig, P. R., Post, D. L., Reis, G. A., & Simpson, M. A. (2003). Human factors requirements of helmet trackers for HMDs. *Proceedings of the SPIE Helmet- and Head-Mounted Displays VIII: Technologies and Applications Conference*, pp. 95-103, Vol. 5079, Bellingham, WA: The International Society for Optical Engineering.

Mejdal, S. McCauley, M. E., & Beringer, D. B. (2001). *Human factors design guidelines for multifunction displays* (DOT/FAA/AM-01/17). Washington, DC: Office of Aerospace Medicine, 2001.

Piccione, D, Krebs, W. K., Warren, P, & Driggers, R. G. (2002). Electro-optic sensors to aid tower air traffic controllers. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, pp. 51-55, Santa Monica, CA: Factors and Ergonomics Society.

Rolland, J. P., & Fuchs, H. (2001). Optical versus video see-through head-mounted displays. In W. Barfield, & T. Caudell, (Eds.). *Fundamentals of Wearable Computers and Augmented Reality.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ruffner, J. W., Deaver, D. M., & Henry, D. J. (2003). Requirements analysis for an enhanced vision system for air traffic control tower surface surveillance. *Proceedings of the SPIE Enhanced and Synthetic Vision Conference*, Vol. 5081, pp. 124-135, Bellingham, WA: The International Society for Optical Engineering.

Ruffner, J. E., Fulbrook, J. E., and Foglia, M. (2004). Near-to-eye display concepts for air traffic tower controllers. *Proceedings of the SPIE Head- and Helmet-Mounted Display Technologies and Applications*, Vol. 5442, pp. 120-131. Bellingham, WA: The International Society for Optical Engineering.

Yeh, M. & Wickens, C. D. (2001). Display signaling in augmented reality: The effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors.* Vol. 43, pp. 355-365. Santa Monica, CA: Human Factors and Ergonomics Society.

INTERPRETING CULTURAL AND PSYCHOLOGICAL DIFFERENCES BETWEEN USA AND EASTERN COUNTRIES

Dott. Ing. Giorgio Sacco
Rome, Italy

The interpretation and explanation of mainly Western cultures with reference to family configuration may help developing theory and practice concerning mistakes and errors, right decisions and good choices.

Introduction

Purpose and Scope

A purpose of present article (see also G. Sacco, 2003) is to briefly illustrate the different kinds of family configuration, and their involvement in personality and culture development. The inherent most direct concern is search for right behavior and good choices, plus avoidance or management of errors and mistakes respectively, with special reference to aviation safety. There are signs that improvements in such field would be worth wishing for, see e.g. S. Dekker (2003), and this article is aimed at making a step towards such improvements.

Methodological Considerations

Quite coherently with the civil aviation context the address of present article is multi/intercultural and multi/interdisciplinary, in other words is considering both differentiation and integration between cultures and between scientific-professional disciplines, as applicable. Within that address a methodological aspect is using further approximations. G. Hofstede (2001: 177-179, 181), while proposing a polarization between grand-theory model and empiricism, attributes the grand theory to the High Uncertainty Avoidance Index (UAI) values, unlike empiricism. Beyond Hofstede's evaluations the theoretical certainties of grand theory would address to think that deterministic reasoning would be more properly assigned to High UAI cultures. An example may be the case of the Japanese, where the conduct of each person must be foreseeable. Such divergence between theory and practice is however not without some inconvenience: Grand theory, not doing deviations, from principles and consequences, may fail as correspondence to facts, while the empiricism scheme may fail as grasping here and there, avoiding to make preliminary assumptions. In any case both grand theory and empiricism could have their induction and deduction aspects. With that perspective e.g. the fact that Gauss, the great mind of the probabilistic calculations, was borne in Austria, a High-UAI country within the Germanic Language group, seems to have a sense: his combinatory calculation theory would be a sort of grand theory, but made with the calculation instrument typical of the cultures liking stochastic reasoning. Centuries before that, Galilei was successful by experimentally observing phenomena which today are put in an easy deductive form. Both together, then, grand theory and empiricism could lead to the best results in scientific terms when a satisfactory agreement between deductive theory and facts may be reached. Just as a hint, comparisons between the (especially "soft") scientific elaboration by different cultures would lead to more complete scientific views. Even Hofstede's (id.,: 177-79, 53-56) work, based on an empiric statistical inquiry (evidencing the dimensions Individualism and Masculinity) and two theory-driven statistical inquiries, for UA and PD (Power Distance, that is the hierarchical distance perceived as existing between a powerful person of a Company and the employees), keeps into account both the above possibilities, principle-based theories (however statistically inquired) and empiric statistics, plus their integration. A composite methodology is therefore deemed necessary and opportune to deal with the subjects included in the present article.

Deep Down Terminology

Words and their definition, meaning, are the necessary elementary tools of this and any exercise of verbal logic. Let's try to consider here even their deepest and most ancient features.

Anthropology-based Terminology

Connected directly to psychological considerations may be the distinction between tendencies to generalization and differentiation with reference to the terminology for agnates and affines (see G.P. Murdock, 1949, Ch. 7). Roughly, a distinction can be made between: terms based on identity (e.g. father's brother called even father), which would express an ethnic (mechanical) solidarity, and terms based on differentiation, which would be more inherent to marriage, civil life. They would also respectively be inherent to a conduct following the customs line, and to a "normal" conquest of the "object", that is, which doesn't jeopardize the needed solidarity above. A note may be that, while in the contemporary Western world there is a general diffusion of terminology based on differentiation, examples and cases which could virtually be represented by terminology based on identity may still exist. Let's think to the psychological growth and forming: such characteristics could be useful in determining the possible kinds of a child's psychological growth process within his family, if practically ending with an identification with other significant people or institutional structures, or in a readiness to common work, or in a more or less happy compromise between these two extremes.

Application to errors and mistakes. Mistake and error could then be two aspects of the same wrong, abnormal fact, whose corresponding right, normal reference would be typical of a certain culture. Therefore it could be not much proper or useful trying to determine mistakes and errors independently from a defined set of rights and laws, a certain culture. However they can be included in a comparison between cultural traits and characteristics of different cultures. While doing that a notion of "composite systems", e.g. systems composed by a "mechanical" and an "organic" part, could be found suitable and useful.

Field-dependence and not. Witkin did introduce the notion of field dependence (see e.g. Okonji,1980, Nisbett et Al.,2001), where field independence would correspond to the above mentioned differentiation, to a more articulated personality. Field dependence instead would be more archaic, and would have also some feminine characteristics. Anyway, being such aspects applicable to the above described fundamental language characteristics, it should be possible to conjugate them conforming to the many different Murdock's types of primitive social structure. In a culture more based on identification ties, and therefore on generalization, there would be more tendency to neglecting empirical aspects. On the contrary in a culture much using differentiation there would be a tendency to avoiding theorization. This second category of facts would be the most promising for the evolution into a civility, even with the defect of avoiding principles and their consequent grand theories, and also if corresponding to migrants. In such kind of society, mostly based on contracts concerning "external objects", venalities, a definition of Mistake with reference to contracts stating the right "takes" seems to be quite acceptable. The same way blames, as more inherent to generalization, would be su bject to be possibly avoided. Hence would derive the possibility to build entire organizational systems based on such aspects: contracts, empiricism, avoidance of grand theories and of blames. A different if not opposite way was followed by ancient China: in fact Confucius was successful in changing customs of his society from external to internal values, from guilt to shame. An additional contribution in that sense was that of the Yin-Yang philosophy. At the level of stress elaboration something similar appears: in Low IDV cultures (see e.g. many Oriental countries) people are more self-adaptive, in High IDV cultures people are more seeking to change the environment, (from Olah, and from Essau & Trummensdorf, both cited in Hofstede, 2001: 242, 518). Accordingly, while Oriental religions are greatly reciprocally tolerant, Western monotheistic religions are much intolerant. Even speaking about internalization (see Lynn & Hampson, cit. in Hofs.: 188), introversion appears to be correlated to High PDI, with an implication of High UAI. At that point it appears opportune to note that field-dependence, conjugated as above, would encourage to see similarities between at least High PD, UA, Collectivism, as all referable to different identification structures, but similar as elementary identification mechanism. Independence, search for change, individualism, would even be someway related each other.

Error and Fall. Known sentences by Cicero, St. Augustin, Benjamin Franklin, distinguish two degrees in Error: one, quite easy to be forgiven, the other, concerning persistence in error, would imply social degradation, rejection. This difference remembers a bit that between the different phases of Selye's General Adaptation Syndrome (GAS): quick correction or chronic incapability to recover from certain kinds of error. Where impossible to recover from errors a condition of resistance would succeed. That is quite clear even at the social level, in the hierarchy of social classes defined by the Christian Church in the middle Age. Of course the GAS has its particular cultural fields of applicability. Just to have some reference, Neuroticism is correlated mostly to High UAI, see Lynn et al., cited in Hofstede, 2001: 155-57, 514-15; conversely, there would be a positive correlation of psychosis incidence with Low UAI.

The above further way of considering the error would be quite in agreement with the aeromedical perspective illustrated by Wiegmann and Shappell (2001), which would be favorable to the possible consideration of fatigue and stress as causes of accidents. It also would suggest a comparison beyond the limits of Western culture, with systems like the Indian stratification into different Castes, or with the above mentioned Oriental characteristics.

Western facts. The health-stress-morale complex has been found influencing the rate of maintenance errors (Fogarty, 2001). Aircraft pilots' stress and fatigue have been the subject of thorough campaigns in USA for the reduction of pilots' flight hours.

Towards a definition of error. A fully satisfactory definition of error would be still not stated, see D. Wiegmann and S. Shappell (2001), in addition to S. Dekker (2003). Interior forms of error are deliberately avoided (Hollnagel, 1998: 26). There is no need to go beyond the limits of the Western Civility to say something more on that subject: in fact the definition of a contractual systemic world is already an effort against certain characteristics of the Western culture itself, e.g. Human Relations and further developments in the same sense against Taylorism. Let's however return later to the subject of errors and mistakes.

Families Throughout the World

People normally grow in families, and this may shape their minds. Family kinds may then be a reference for both cultural and psychological features.

Stem Families and Their Derivations

With reference to the contemporary Germanic Language group, to which belong populations of partly common origins, the generation of social differences in the last centuries may be at least partly referred to the stem family and absolute nuclear family (i. e. the nuclear family based on inequality) dynamics (for such names and other surrounding notions see E. Todd, 1983, 1990). In fact the connection between those family systems, both based on inequality among brothers, and the inherent social system appears to be quite direct. The work of F. Sulloway (1996) for the contemporary Western brothers' groups may give an idea of the inequalities between brothers, which appear to be determined spontaneously, as a function of birth order only. That would correspond also to differentiation between social categories, and, as a likely hypothesis, even between entire populations. As of the countries belonging to the Germanic languages group a certain comparison on family structure may be done between it and the Slavic populations. The traditional Slavic family would include quite matriarchal aspects (see E. Gasparini, 1973). The initial regime of the Germanic populations seems to be similar to the quite egalitarian Slavic regime, see Laura Thompson (1969, Ch. 10), E. Gasparini (1973: 267), and the custom of the Borough English (see on the other side the case of Slovaks). Franks did introduce primogeniture and Feudalism. In at least some Nordic countries the passage to Christian religion was done by the creation of christianized kings, who by the Church's consensus and authority would have strengthened their position (see some example in G. Jones, 1968, 1973). On the other side the personal characteristics of the Protestant religion did favor even migrants, in this case favoring the heads of each single family. The inclination of the monotheistic Western religions to Manichaean distinctions and externalization of the Evil with respect to the Good did probably favor the creation and consolidation of a partial fracture between the American Anglos and many of their European relatives. In sum, the overlap of Christian religion and its annexes to other more ancient cultures did often lead to noticeably peculiar compromises which don't help at all in distinguishing e.g. between error and mistake, in the sense shown above. Nevertheless the consideration of this maze and its derivations may help in understanding and explaining many single aspects of contemporary Western culture. Much clearer is the stem family system of the Oriental countries, to which corresponds a characteristic decrease of UAI with the increase of PDI and 100-IDV (here indicated as COLL). A hypothesis on it may be that there is a sort of potential competition between the father–heir alliance and a potential group of brothers. The underlying economical scheme is agricultural. Hints in favor of a similar transversal growth (UAI normally grows with PDI) appear to exist among some countries of the Germanic language group, however in the opposite sense with respect to a central strip including the other populations in a PDIxUAI plot (see Hofstede, 2001: 152) .

Organizations. At the level of industrial activities an example of religious influence is Taylorism, one of whose organizational features was avoidance of errors at the level of common workers. The further intervention of remedies in the Human Relations' sense appears to be consistent with the consideration of motivation, and then choices, mistakes. Probably the differentiation between Direction and Management, Decision and Choice, may correspond to such different and complementary tendencies, especially in the USA. Blame to the individual or criticism to the system, a not much dissimilar polarization which has signed the recent history of thought on human error, appears to be consistent with PD (Hofs.: 97, 98), in the sense that a low PDI would be favorable to a criticism towards the system (see also on Organization Development, Hofstede, 2001: 390). In other words blame to individual could be connected to field dependence, but as a rejection from a field-dependent group, as for a sort of scapegoat, or reality denial. A note concerning PD is that USA PDI value is relatively high with respect to some European countries within the Germanic Language group: that would be in agreement with a greater religiosity of the USA (see Inglehart, 1997, fig. 3.3), and therefore with the said spirit of Taylorism. The history of such questions would be also in agreement with the fact that the technological design would still be the hard, "deterministic", grandtheory-like nucleus, and other more choice-oriented considerations would be the surrounding part. So it is also for the UK socio-technical theory. High expectations on the benefits of technology result to be correlated mostly to High PDI (and quite high UAI), and negatively to high IDV, even if technology is more used in lower PDI societies (Hofstede, 2001: 101, 107, 506, citing Inglehart). The same is for automation: as R. Helmreich and A. Merritt (1997: 97) refer, pilots who like/ prefer automation correspond to high PDI, quite high UAI values, and negatively to IDV values. This would lead to think that an aspect of technology would be a sort of combine involving "mechanical" aspects, even in a cultural sense, and that consistent with it would be error, in the mechanical sense, not mistake.

Juridical systems and contracts. Some comparison between juridical systems (see Hampden-Turner & Trompenaars, 1997: Ch. 8; Hofstede, 2001: 174, 180-81, 505), or anyway considerations concerning rights (Trompenaars, 1998: Ch.. 4) may also be hinted. They may also be done in a way quite parallel to the distinction between Error and Mistake. That is, there would be systems based mainly on inter-individual object-based contracts (that is lacking of emphasis on genetic similarities) and systems based more largely on ethnic solidarity (e.g. the brotherhood-based gentilitial Latin system, and the unilineal descent characteristics of the Chinese complex). The advent of civilization has led to a strong increase of sub-systems and regulations based on the above contracts, however they are normally part of systems which include at least small nuclei of ethnic solidarity. But it is probably the case to illustrate how different can be the Latin and the German models of law. A quite high UAI value of Germany is not to be deemed equivalent to the typical Latin one. In fact the German model of law is known as being mostly prohibitive, unlike the English one. An interpretation of this fact may be that one of its aims would be that of creating and conserving differences between the roles of heirs and non-heirs, and possibly between the corresponding different social classes which may be generated by that. A concern for errors and mistakes is: while speaking about rule-based behavior, about which rules is one speaking? More widely, one should better clarify what would be meant when speaking about "familiarity" for the SRK taxonomy.

Another consideration, which would become more evident just by speaking about legal systems, is that the comparison between Eastern countries and USA is not so easy. Doing that on the basis of a cultural dimension alone, the Individualism (IDV), appears to be not enough. The characteristic favorable inclination towards Rules and Categories of the ancient Greeks and Romans (see e.g. Nisbett, R. et Al., 2001) would appear to be referred to another cultural dimension, UA, which would imply the above differences in law structure. China's equivalent of law would be referred directly to a sort of almost religious, knowledge-based power. PD and UA in fact appear to be referable to generalized Knowledge and Rule dimensions respectively.

The "East to West" Composite Scheme

Hofstede's (2001: 152) UAI x PDI plot (see also Fig. 1) is taken as reference for the representation on a plane. In it a main area appears to be made by the two

High PDI	Low PDI	
<u>Stem family with Low IDV</u> *Excessive dependence, introversion* China	<u>Absolute nuclear family</u> *Mistake, lack of theory* Denmark, Sweden Ireland, UK, USA	Low UAI
<u>Egalitarian nuclear family</u> *Scarce consideration of experience* Romans, Greece	<u>Stem family with High IDV</u> *Neuroticism, GAS* Germany, Austria.	High UAI

Table 1. *UAI x PDI plot (schematic)*

sections Low PDI Low UAI and High PDI High UAI. With respect to it the other two sections would appear lateral and less balanced, and a minor number of countries is found in them. The High PDI Low UAI (and Low IDV) section would be characterized by high sensitivity, low level of activity, while on the opposite the Low PDI High UAI (and High IDV) section would be characterized by high activity, e.g. wars. The most distinguishing characteristic of the plot appears to be PD. To its stripe would appear to be more properly connected the word Family. The High PDI sections would address to an interesting comparison between China and Ancient Romans, the power of the single governor against the "law equal for all" model of the Romans (see in Hofstede, 2001: 181). The low PDI sections are occupied almost entirely by the Germanic language populations. In them two different tendencies, towards low and high MAS (Masculinity), are quite intermixed, but still distinguishable. They are instead well distinguished on a IDV x MAS plot, Tab. 2 (see also Hofstede, 2001: 294), while the Low IDV area is not so much polarized by MAS. In it Low MAS countries may be easily attributed to the Nordic culture, while high MAS countries are placed in the geographic areas previously occupied by the Celts, whose culture is possibly partly still living. Different original mythologies (lunar male and prevalence of number three, solar male and prevalence of number four) would correspond to the different MAS values. A possible hypothesis about would be that, while the Lunar cultures would be originally more inherent to agriculture, the Solar cultures would be inherent to herders, at least within the Germanic Language group. This distinction is valid also for the USA, where in the north there would be more peasants of UK, Dutch and German origin and in the South more herders, of Irish and Scottish origin, see Nisbett, R. and Cohen, D., 1996. The aim of these considerations would be the individuation of different family configurations, for the concern of the attribution of characteristic errors and mistakes to them. In Tab. 1 are added hints on the most significant family kinds, and on possible or real kinds of mistake/error.

As hinted above families would be at the confluence between National and psychological characteristics.

	Low MAS	High MAS
Low IDV	Portugal	*High traffic deaths* China
High IDV	<u>Nordic culture</u> Denmark, Norway, Sweden	<u>Celtic tendencies</u> *Stress, burnout* Austria, Germany Ireland, UK, USA

Table 2. *IDV x MAS plot (schematic)*

E.g mistakes could correspond to an externalization of psychological facts which in other cultures would be more unconscious, a known fact for e.g. the USA. But the clearest one is about stem families: in the Chinese culture by Low IDV there would be harmonization between male and female, possibly both peasants at the origin, while in a soldier-based configuration one could conflict with own parents, especially father. A similar condition would exist in Japan, for the Samurai. In the Western countries it is traditionally associated with Error, in the Biblic sense of being expelled from the Eden and searching for a new place. Its place in Table 1 would be in the High UAI Low PDI section, quite coherently together with Neuroticism. On the opposite side, High PDI Low UAI quadrant, the inherent characteristics should be: staying resigned and peaceful with own father, ignoring the temptations of too incongruous external objects. Partly similar features exist even in the Western society: Smith (1986) found a tendency to vertical ordering in the civil society (High PDI), and lateral in the aristocratic and military society (High UAI, however distinguishing the cases of the Latin and Germanic language groups).

Possible integration. In many Eastern cultures a double mindset does exist: individualistic people and not, externalized yin-yang and not. That probably corresponds to the presence of both heirs and not heirs in the same region. In other words, while, due to migration in USA and Canada, a stronger division does exist between Anglos and e.g. Germany, in those Eastern countries the two possibilities are more reciprocally integrated. Aviation, together with the development of other communication means, should favor a better integration between those elements, whose separation is enhanced by some characteristics of the Christian religion (remember also the above subsection "*Field-dependence and not* ").

Mistakes, Errors and So On

Accident Data

Low IDV, in relation to High UAI and MAS, corresponds to high traffic deaths. Notwithstanding a quite high IDV Austria has the highest traffic death rate (Id: 199, 243, on United Nations' data, 1973, concerning 14 European Nations). That would confirm some hypotheses on the High UAI Low PDI section. In addition High MAS would be correlated to high stress and burnout (Id.: 316, 318, citing Schaufeli and Van Dierendonk), and that would contribute to explain the above data on Austria. Soeters & Boer (2000) found a correlation of European military aviation accidents also with High UAI. Also Lynn & Hampson (cited in Hofstede, 2001: 156, 188) list a high accident death rate as a component of a "neuroticism factor" correlated with High UAI. On the opposite side high civil aviation accident rates were found related to High PDI, Low IDV(the last overwhelmed by Low GNP) (Weener & Russell, Ramsden, see for both in Hofstede: 131, 115). The exemplar case would be that of many Oriental countries, see e.g. H-S Jing (2002). That would be very good for the following subsection, however obviously stronger confirmation would be needed.

Perrow. A comparison with Perrow's theory seems at this point almost unavoidable, at least as a hint. In fact UA should be comparable to Perrow's fixedness of the elements, and PD, that is the degree of obedience, to the linearity of action following a command, and also to the degree of centralization. Two cases would appear more critical, of an excess of PD with respect to UA and vice versa. In the other two cases the existence of a nuclear family would warrant a greater equilibrium. The existence of at least two other Hofstede's dimensions would show the limits of C. Perrow's theory.

Error. It seems now possible to give at least a more complete description of error, that is including the socially-relevant systematic aspects, a continuous state in error. That would include many forms of divine punishment which may be encountered in histories and mythologies. Especially under this form the error would be mainly a consequence, typically the consequence of a mistake in a culture which would someway foresee it and its consequences. Typical would be the wandering consequent to a mistake, a sort of exile. However migration in many cultures wouldn't correspond to unsustainable mistakes, and also often the migrants are more lucky than the heirs. But in some cases a deep state of error would be the consequence of a very heavy guilt. In those cases a full definition of error wouldn't be recommended, because the most exemplar cases would risk to be even the most unbalanced, which couldn't be a good example for other aspects.

Concluding Remarks

It isn't scope of this article to define that some cultures would be better than others, at least from the viewpoint of errors, mistakes, safety. However it seems that on the basis of the above notes, especially those concerning the comparison with C. Perrow's theory, some favor may be given to the cultures having more confidence with some form of nuclear family. In fact, at least on the basis of qualitative considerations, this would be a warranty of a better psychological and social equilibrium. From the statistical viewpoint a note is spent in favor of the less rigid cultural forms, that is corresponding to high values of IDV, low values of PDI, UAI. However it is possible that this is a result of a transient cultural situation, and that new tendencies towards different ways to consider science and technology may lead to further and different contributions to safety and safety culture. Throughout the article many important points have been touched: on the nature of science, on mistakes and errors, on different cultures for the concern of safety. None of them has been dealt with to arrive at specific final conclusions, also for the intrinsic limits of the kind of article, and for the possibilities of subsequent further developments.

Bibliography

Dekker, S. (2003) *Illusions of Explanation: A Critical Essay on Error Classification*. Int. J. Av. Psych. Vol. 13, No. 2, Pages 95-106.

Fogarty, G.J. (2001) *The role of individual and organisational factors in aviation maintenance*. 11th ISAP Proceedings.

Gasparini, E. (1973) *Il matriarcato slavo*. Sansoni .

Hampden-Turner, C. & Trompenaars, F. (1997) *Mastering the infinite game*. Capstone.

Helmreich, R. and Merritt, A. (1997) *Culture at Work in Aviation and Medicine*. Ashgate.

Hofstede, G. (2001) *Culture's Consequences*. Sage.

Hollnagel, E. (1998) *Cognitive Reliability and Error Analysis Method*. CREAM. Elsevier.

Inglehart, R. (1997) *Modernization and post-modernization: Cultural, economic, and political change in 43 societies.* Princeton University Press. Princeton, NJ

Jing, H.-S. (2002) *The Dragon in the Cockpit*. Human Factors and Aerospace Safety 2(3), 257-275

Jones, G. (1968, 1973) *A History of the Vikings.* Oxford University Press.

Murdock, G. P. (1949) *Social Structure*. The Macmillan Co., NY.

Nisbett, R. et Al. (2001) *Culture and systems of thought: holistic vs. analytic cognition*. Psychological Review, 108, 291-310

Nisbett, R. and Cohen, D. (1996) *Culture of Honor: The Psychology of Violence in the South*. Westview Press.

Okonji, M. O. (1980) *Cognitive styles across cultures.* In: N. Warren (Ed.), *Studies in Cross-cultural psychology*. Vol. 2. Academic Press.

Sacco, G. (2003) *Errors, Mistakes, Cultures.* Proceedings of the 12th ISAP, Dayton, OH.

Smith, A. (1986) *The ethnic origins of nations*. Basil Blackwell, Oxford.

Soeters, J. & Boer, P. (2000) *Culture and Flight Safety in Military Aviation*. IJAP 10(2),111-33.

Sulloway, F. (Ed.) (1996) *Born to Rebel*. Random House.

Thompson, Laura (1969) *The secret of culture. Nine community studies*. Random House.

Todd, E. (1983) *La diversité du monde*. Seuil.

Todd, E. (1990) *L'invention de l'Europe.* Seuil.

Trompenaars, F and. Hampden-Turner, C. (1998) *Riding the waves of culture*. McGraw-Hill.

Wiegmann, D. and Shappell, S. (2001) *Human error perspectives in aviation*. International Journal of Aviation Psychology 11(4), 341-357.

HUMAN FACTORS DESIGN OF ELECTRONIC DOCUMENTS

Thomas L. Seamster
Cognitive & Human Factors
Santa Fe, New Mexico

Barbara G. Kanki
NASA Ames Research Center
Moffett Field, California

The Federal Aviation Administration (FAA), working with the Master Minimum Equipment List (MMEL) Industry Group, is developing a new MMEL electronic format. The MMEL refers to a series of documents controlled by the FAA that lists equipment that may be inoperative under certain conditions while still allowing the aircraft to be airworthy. Each aircraft model has an MMEL, and operators must work with that master document to determine the relief items for their specific aircraft. The resulting Minimum Equipment List (MEL) for an operator's aircraft is used by both ground personnel and pilots to determine the procedures for maintaining airworthiness. Currently, the MMEL is available in text format, and the industry needs an electronic format that is more efficient and that will be compatible with key aspects of future data standards. Members of the MMEL Industry Group were surveyed to determine the main user needs and human factors considerations for the development and evaluation of the MMEL electronic format. This study identifies key operator needs that can direct the development of not only the new MMEL format but also the broader category of aviation electronic documents.

Introduction

As the aviation industry continues the transition from paper to electronic documentation, the opportunity for standard and more efficient information exchange and reuse has been recognized, but the emphasis has often been on technical and engineering solutions and human factors efforts have not given full consideration to the needs of the different user groups. This study provides an opportunity to understand operator electronic document needs encompassing ground and flight organizations within airlines.

Organizations including the Air Transport Association (ATA) and the Federal Aviation Administration (FAA), with help of operators and suppliers, have been considering different approaches to the structure and format of aviation information used in flight and maintenance operations. There is a growing consensus that current technology, such as the eXtensible Markup Language (XML), can offer a viable solution, but less progress has been made on defining a complete set of user requirements and human factors considerations that should form the basis of electronic format, specifications or standards efforts.

Recently, the Master Minimum Equipment List (MMEL) Industry Group (IG) in collaboration with the FAA have been looking at ways to select or develop an MMEL electronic format that will facilitate MMEL publication and revision. The MMEL refers to a series of documents controlled by the FAA that lists equipment that may be inoperative under certain conditions while still allowing the aircraft to be airworthy (FAR 121.628). Each aircraft model has an MMEL, and operators must work with that master document to determine the relief items for their specific aircraft. The resulting Minimum Equipment List (MEL) for an operator's aircraft is used by both ground personnel and pilots to determine the procedures for maintaining airworthiness.

This is a strategic project because enhancements to the MMEL revision process will, in turn, improve the MELs, a key document used by maintenance, dispatch, engineering and other ground personnel, as well as flightdeck crews. Further, in their timing, the MMEL IG is in a lead position to establish a working electronic format that can influence future data standards in related areas across the aviation industry. Because of the potential improvements and influence on aviation electronic documents, the MMEL IG was surveyed to determine key factors that should be considered in the development of a MMEL electronic format. The results are analyzed in the context of authoring, revising and reusing aviation information in more standard and efficient ways across ground and flight operations.

Background

The NASA/FAA Operating Documents Group has been meeting as an industry group over the past eight years to address key operating data and document issues (Kanki, Seamster, Lopez, Thomas, & LeRoy, 1999; NASA/FAA, 2000). With the shift from documents to electronic data, the Group has focused on user requirements that should be addressed during this significant transition. Although the NASA/FAA Group has identified a wide range of issues including safety critical data, standardization, human factors, and security (Seamster & Kanki, 2002), the emphasis here is on human factors and end user requirements.

When working with industry data requirements, there are two important dimensions to consider. First is the interaction between operator, supplier and regulator requirements. Traditionally, each group has concentrated more on their own requirements and less on the overall industry needs. The challenge is to recognize the differences and define the common ground that can be used to develop industry requirements. Operators are most interested in the efficient conversion of supplier documents into their own document formats. Suppliers tend to concentrate on the efficient and accurate production of documents in whatever form the different operators require with less emphasis on standards. Regulators have been more concerned with the approval process, often focusing at the document page level.

The second dimension is the process to product human factor (Seamster & St. Peter, 2002). This dimension highlights the different requirements of those who work with the final product, such as the pilots, mechanics and other end users, compared with those who manage documents and data, such as those in publications. The transition to electronic documents provides an opportunity to ensure that improvements are made for both user groups, the data end users as well as the data managers.

More of the human factors research and guidance has been offered from the end user perspective. Pilots use operational data on the flight deck with its workload management and safety-critical requirements. Existing guidance for electronic documents such as the electronic flight bag (EFB) concentrate on how the system interacts with crews on the flight deck (see Chandra, Yeh, Riley, & Mangold, 2003). Maintenance workers are an important second set of end users with a different set of usability issues as identified by Chaparro, A., Groff, L. S., Chaparro, B. S., and Scarlett, D. (2002). Further, human factors issues related to maintenance documents and procedures including the MEL have been identified by analyzing incident reports submitted to the NASA Aviation Safety Reporting System (see Munro & Kanki, 2003; Patankar, Lattanzio, & Kanki, 2004).
Less research and guidance is available for the data managers and related document developers. Data managers have workflow requirements to automate and simplify the creation, review, approval and distribution of operational information. They share some needs with end users but have additional needs brought on by the data revision and the publication process.

Usability issues from these two types of user groups (end users and data managers) must be considered jointly in such a way that electronic documents and data can be efficiently managed while meeting the safety-critical end user requirements. Both user groups must ultimately participate in developing new electronic formats and standards working with suppliers and regulators. It is important that industry not gain efficiency for one group at the expense of the other.

The MMEL Industry Group has been working to develop a MMEL format that all data management users can access. This new format should allow documents to be electronically accessed and interchanged. XML is a candidate technology for the MMEL format based on the use of schemas. The MMEL Industry Group is tasked to define the requirements with an emphasis on technical capabilities such as the tracking of changes, deleted items and managing effectivity.

The NASA/FAA Operating Documents Group has supported the MMEL effort by identifying high-level human factors considerations concentrating on data management user groups. The reason for this data management user perspective is that it has received less attention up to this point. Moving forward, this perspective along with the results reported here need to be merged with end user requirements across operators, suppliers, and regulators to ensure a format that is usable for the larger aviation industry.

Methods

The MMEL and MEL Usability form was prepared by the NASA/FAA Operating Documents Group in conjunction with several members of the MMEL Industry Group. The purpose of the form is to determine the most important usability requirements as the MMEL Industry Group develops a new MMEL format based on eXtensible Markup Language (XML) schema. The form was designed to help the MMEL IG determine the most important MMEL improvements as they review options for the new MMEL XML format.

The instrument is a one page rating form. The top part of the form asks each rater for their background information to determine their current job, their experience with documents and publications along with their years of MEL experience. The middle section of the form collects data on each rater's perspective on the MMEL revision process. The last section of the form, the focus of this paper, presents 22 possible MMEL improvements asking participants to rate each on its degree of importance using a five-point scale. For these ratings, "Most Important" is given a value of 1 and "Not at All Important" is given a value of 5.

The instrument was administered to approximately 40 participants at a recent quarterly MMEL Industry Group meeting. Some of the participants represented the same organization and worked together on a single form. The group received a detailed explanation of the form and respondents were asked to consider their organization's priorities when providing their ratings. The completed forms were sent on to the NASA/FAA Operating Documents Group facilitators for compilation and reporting.

Results

A total of 28 MMEL IG members completed the forms. One participant provided ratings on less than half of the MMEL improvement items so that data was removed from the analysis resulting in a total of 27 respondents. Of those, 21 represented operators including majors, regionals, and cargo. The remaining six represented suppliers and regulators. Most of these participants were experienced with the MMEL and MEL process having worked an average of 11 years on aviation documents or publications (range from 2 to 24 years), and they also had 11 years of MEL experience with a range from 1 to 23 years.

In addition to the ratings data, degree of certainty was also collected using a three-point scale from High degree of certainty to Low degree of certainty. The extra data was recorded in order to compute weighted scores that would highlight those ratings made with a High degree of certainty over those ratings made with less certainty. Ratings and their means were calculated using both the raw scores and the weighted scores. The results were similar, and because of some missing certainty data, the raw score rating data was used for this paper.

Table 1 shows the MMEL improvements listed in order from most to least important based on the means of the raw ratings. The top six items provide a coherent set of priorities around a more expedited and standard MMEL process involving the identification and authoring of internal MEL revisions. The next group of important requirements support those first six in that they address authoring, revisions, standards and MMEL format.

Table 1. *MMEL Improvements in Order of Importance Based on Ratings of 27 Participants* (1 = Most Important and 5 = Not at All Important)

MMEL Improvements	Rating Mean
Expedited MMEL authorization process	1.68
Identifying MMEL changes impacting your MEL	1.74
Downloading as XML file	2.00
Working with MMEL revisions	2.00
Authoring internal MEL revisions	2.04
Enforcing standard MMEL structure (FAA/manufacturers)	2.04
Identifying all MMEL changes since last revision	2.11
Enforcing standard MPM/DDG structure (manufacturers)	2.11
Reformatting the MMEL for MEL authoring	2.22
Viewing the MMEL in a more usable format	2.22
Improved FOEB process	2.33
Working with upgradable schema	2.35
Tracking effectivity	2.46
Standardizing on one ATA numbering system 6 digits	2.57
Downloading as DOC file	2.62
Handling MMEL deleted items (what was deleted)	2.63
Supporting PDF output for MMEL	2.65
Supporting MS Word output	2.69
Viewing the MMEL in a common browser	2.77
Printing the MMEL	2.93
Downloading as PDF file	2.96
Downloading as EXE file	3.63

The last group with decreasing importance include some of the technical items that are the current focus of the MMEL Industry Group. Supporting MS Word and PDF formats and viewing the MMEL in a common browser are less important, and downloading as PDF or EXE file along with printing the MMEL are the least important improvements.

The most important improvement is an expedited MMEL authorization process. Figure 1 shows that the majority rated it as "Most Important" with the majority of the rest rating it at "Very Important." Independent of the type of technology or format, participants want whatever system is implemented to speed up the MMEL authorization process.

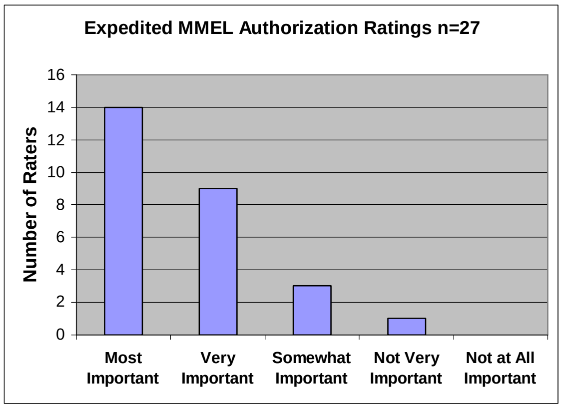


Figure 1. *Ratings distribution for "Expedited MMEL authorization process."*

Specification of the type of technology to be used, in this case XML, was also rated toward the top, but it is interesting to note that the rating distribution was not as clear cut with more participants giving it a neutral rating (Somewhat Important) than those who gave it a "Very Important" rating (see Figure 2).

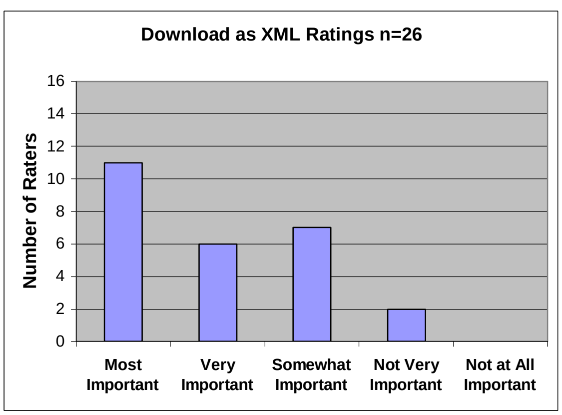


Figure 2. *Ratings distribution for "Downloading as XML file."*

The results of this survey can be very helpful in directing the development of a new MMEL format to be used by the aviation industry in updating their own MELs. The most surprising finding is that this group of data managers are most concerned with general, high-level improvements and less focused on the lower-level technical issues that seem central to current industry efforts. This is significant from a human factors perspective and argues for a greater understanding of the MMEL and MEL authorization and revision process based on a user-centered rather than a technology-centered approach.

Discussion

The results of this survey emphasize the relative lack of attention paid to the requirements of the data manager user group. Even as pressure is applied to the document developers and distributors for timely, accurate information, the data managers work within a system that is inefficient. The top-rated improvements requested by this group of users are relatively high level process oriented changes that can improve both accuracy and efficiency. More specific, technical improvements are valuable but in themselves, fail to set the system-level efficiencies and standards that are needed.

The process and results reported here suggest ways to improve the development of electronic data formats and to foster the acceptance of the resulting formats and standards. Interpretation of these results indicates several ways to improve industry participation and also suggest ways to improve the acceptance and successful implementation of the resulting standards.

To achieve industry acceptance, a user-centered approach must be used in the development of electronic data solutions. As the technical work proceeds in developing electronic formats for the MMEL and other documents affecting flight and maintenance documentation, it is essential to continue collecting data and working with key user groups such as the flight operations data managers and the flight deck data end users. In the case of the MMEL and resulting MELs, a large number of users will have a stake in the process. But as indicated by the ratings, it may be more important to address the industry-level requirements and standards first since these improvements can alleviate the inconsistencies that data managers must work around.

Similarly, the consistencies and standards developed at the company level can alleviate the inconsistencies that individual end users across the company must work around. For example, when complementary procedures are developed or updated independently within their own organizations (e.g., pilot and ramp de-icing procedures), inconsistencies can develop because each group has its own tasks, responsibilities and priorities. Even the most basic terminology and format differences may go unnoticed for a long time. In contrast, if all company procedures are built and revised from a common reusable data source, this type of divergence can be avoided without compromising end user requirements.

The results of this survey represent a small step in the larger user-centered approach to developing new aviation information data formats and structures. Fortunately, data manager requirements have been identified early in the process providing an opportunity for additional steps that can ensure a good fit between electronic data formats and structures and aviation industry user groups. The next steps include:

- Identify and analyze key data management and authorization tasks most tightly coupled with the MMEL format
- Identify relevant human factors measures to be used in the evaluation of new electronic data formats and structures
- Harmonize data management requirements with end users across operations, suppliers, and regulators.

In summary, the MMEL IG has identified top-rated improvements which focus on the resolution of industry-level processes and standards. While not directly addressing end-user issues, these are fundamental improvements required for a better workflow and greater efficiencies in the preparation and timely distribution of essential information. By providing data managers with these improvements, they can be more responsive to their multiple end users because they have the data structures to support effective and efficient document management.

Acknowledgments

We acknowledge the contribution made by all NASA/FAA Operating Documents Group participants over the past eight years. We take this opportunity to thank them with special recognition to Keith Tollar at FedEx and Mark Lopez at Delta Air Lines for their help on the MMEL survey and data collection. Special recognition goes to Eleana Edens (FAA AFS-230) who has supported the NASA/FAA Operating Documents Group as well as to the NASA Ames Research Center, Human Factors Research & Technology Division.

References

Chandra, D., Yeh, M., Riley, V., & Mangold, S. J. (2003). *Human factors considerations in the design and evaluation of Electronic Flight Bags (EFBs), Version 2.* DOT-VNTSC-FAA-03-07. Cambridge, MA: USDOT Volpe Center.

Chaparro, A., Groff, L. S., Chaparro, B. S., & Scarlett, D. (2002). *Survey of Aviation Technical Manuals Phase 2 Report: User Evaluation of Maintenance Documents*. Washington, DC: Federal Aviation Administration.

Kanki, B. G., Seamster, T. L., Lopez, M., Thomas, R. J., & LeRoy, W. W. (1999). Design and Use of Operating Documents. In *Proceedings of the 10th International Symposium on Aviation Psychology.* Columbus, OH: Ohio State University.

Munro, P. A. & Kanki, B. G. (2003). An Analysis of ASRS Maintenance Reports on the Use of Minimum Equipment Lists. *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 851-856). Dayton, OH: Wright State University.

NASA/FAA (2000). *Developing Operating Documents Manual*. Moffett Field, CA: NASA Ames Research Center.

Patankar, K., Lattanzio, D., & Kanki, B. G. (2004). A Content Analysis of Representative Procedural Errors in ASRS Maintenance Reports. *Proceedings of Safety Across High-Consequence Industries Conference* (pp. 167-172). St. Louis, MO: St. Louis University.

Seamster, T. L., & Kanki, B. G. (2002). Future of Aviation Operational Information. In T. L. Seamster & B. G. Kanki (Eds.), *Aviation Information Management: From Documents to Data* (pp. 163-178). Aldershot, UK: Ashgate Publishing Limited.

Seamster, T. L., & St. Peter, N. (2002). Context of aviation operational information. In T. L. Seamster & B. G. Kanki (Eds.), *Aviation Information Management: From Documents to Data* (pp. 3-18). Aldershot, UK: Ashgate Publishing Limited.

IMPROVING NOVICE FLIGHT PERFORMANCE USING A FUNCTIONAL AVIONICS DISPLAY

Carl F. Smith and **Deborah A. Boehm-Davis**
George Mason University
Fairfax, VA

Supporters of functional interface design argue that direct interaction with the essential functional relationships of a system may aid in the acquisition of domain-specific skill. To evaluate the potential use of a functional display in assisting in the development of piloting skill, twenty novices were trained on either a conventional display or an alternative display that displays the functional relationship of power and airspeed (the Oz display). Novices trained on the functional display showed greater control of power and less deviation from a flight profile over multiple maneuvers. Implications for future research and potential uses in training are discussed.

Introduction

Research suggests that interfaces designed to provide system operators with high-level, perceptual information regarding system properties may improve overall performance (Rasmussen and Vicente, 1992). It is argued that such interfaces should allow for direct perception of the system goal as well as successful performance boundaries (Rasmussen, 1999). One method for accomplishing this may be through displaying information about the abstract, functional relationships occurring within a system. Multiple laboratory studies have shown performance advantages in employing such interfaces (see Vicente, 2002 for a review).

The potential of such functional interfaces has also been noted in the aviation community. Lintern, Waite, and Talleur (1999) argued for cockpit design that allows pilots to directly perceive and interact with essential functional properties of flight, reasoning that direct observation of functional flight relationships may improve pilots' ability to acquire and maintain basic piloting skills.

Limitations in technology and costs associated with implementing and testing functional devices limited empirical evaluations of many functional displays, often reducing the implementation of such designs to proof-of-concept tests (Dinadis and Vicente, 1999). However, progress in display design has led to the implementation of prototype devices that allow for empirical evaluation of the effect of functional displays in an aviation domain.

One alternative cockpit display that displays some functional properties is the Oz system, a graphic interface designed for general aviation (See Figure 1). The Oz display integrates the physical information expressed on a conventional display into a series of basic perceptual forms, creating a display that leverages several emergent feature properties (Bennett & Flach, 1992) to communicate physical and functional flight information. One functional relationship represented by the Oz display is the functional relationship between power and airspeed. A colored vertical line is employed with one color (green) communicating the amount of power being used and another (blue) communicating the amount of power available. The same vertical line's position on a horizontal axis communicates current airspeed. The intersection of the green portion of the vertical line with the angular wings indicates the optimal power setting needed to maintain the current airspeed. Using this graphic, a pilot can directly perceive the most effective and efficient use of power to attain a given airspeed.

The Oz display provides an effective testing ground to examine the effect of functional visualizations in an aviation domain. By comparing performance using a functional (OZ) and conventional display in an aviation task, the effect of employing a functional visualization can be examined.

Smith, Boehm-Davis, and Chong (2004) compared experienced pilots' performance using the OZ system against a conventional general aviation display. Results showed pilots using the Oz system were better able to set and maintain optimal power settings, and showed less deviation from power settings overall. Multiple maneuvers revealed less variability among pilots using the Oz display than those pilots using the conventional display.

The previous findings support the use of a functional display to maintain pilots' current skill set – specifically, the efficient use of power to attain and maintain airspeed. However, these results apply to already knowledgeable experts, and do not directly address the potential for improvements in skill and knowledge acquisition through functional display use.

Figure 1. *The Oz display overlaid on a Cessna 172 in Microsoft Flight Simulator 2002.*

The current research was conducted to examine the effect of a functional display on novice performance and knowledge acquisition. By comparing novices learning to fly using a functional display (Oz) against novices learning to fly using conventional instrumentation, we can evaluate the effectiveness of each interface in supporting novice performance.

If the assumptions of functional interface design are accurate, we would expect to see greater control of power in the Oz display condition than the conventional display condition. A greater control of power would be seen as less deviation from the optimal power setting. Greater understanding of the functional relationships of flight should also lead to greater overall performance, which would be reflected as less deviation from the target flight profile in the functional (Oz) condition.

Method

Participants. Participants consisted of undergraduate students drawn from the George Mason University undergraduate subject pool. Twenty undergraduates (13 males and 7 females) participated, ranging in age from 18-23 years (mean = 20.3 years). None of the participants had any prior flight training or experience. All participants were compensated with class credit for participation. All participants reported normal or corrected-to-normal vision, and reported that they were not colorblind.

Apparatus. An Elite iGATE Personal Computer Aided Training Device (PCATD) driven by a PC running Microsoft Flight Simulator 2002 (MSFS 2002) was employed to simulate the flight environment. MSFS 2002 was configured to simulate a Cessna 172D flying over Dade County Airfield (KDCD). The OZ display was run by the same PC, and covered the central 6 dials of a conventional Cessna instrument panel (see Figure 1).Flight performance data produced by Microsoft Flight Simulator 2002 was broadcast on a local network to another computer for data collection.

Participants were given a demographic questionnaire prior to participation, and then given a packet of slides to follow along with during flight training. Participants were given a paper pretest and a series of questions to answer at the completion of the performance segment of the experiment.

Experimental Design. A repeated-measures mixed design was used, in which display was between subjects, and trials and maneuvers were within subject variables. This yielded a 2 (Conventional/Oz display) X 6 (Trials) design, with 11 maneuvers nested within trials. Trials were administered in three sessions per display, with each session divided into two sets of trials. In the first trial for each of the first two sessions, the novices received feedback from the experimenter during performance. In the third session, the novices received no feedback. Each participant performed the same maneuvers, though the presentation order of maneuvers was counterbalanced across participants. Each participant performed eleven maneuvers per trial (see Table 1).

Procedure. Participants attended a lecture detailing the basic principles of flight and introducing the instrument panel. Principles and maneuvers that were not readily understood by the participants were demonstrated by the flight instructor on the simulator. The time required to complete the training session was approximately ninety minutes.

Following the training session, participants were seated at the simulator and familiarized with its controls. When operating the conventional display, participants were presented with a power table reference for the simulated Cessna.

Participants were then instructed to perform maneuvers by the experimenter, who was seated at a station to the right of the simulator. Participants were given specific instructions on the objective of the maneuver and told to fly each maneuver as accurately as possible. Each maneuver was ended when the participant leveled off within 10 feet of the target altitude and 3 degrees of heading. After a maneuver was completed, the aircraft was adjusted by the experimenter to the position required for the next maneuver.

Participants performed each maneuver for a total of 11 maneuvers, or one trial. After the 11 maneuvers were performed, participants were excused for a short break, and then returned to perform another trial of 11 maneuvers. A set of two trials were considered one session.

For each trial involving experimenter feedback, the experimenter monitored participant performance of the novice and offered guidance based on the principles taught in the instruction session. To ensure *consistency and avoid bias, guidance was limited to a series of phrases directly related to the material initially taught to the novices (See Table 2). After three sessions were completed, the participant was given a document containing a series of open-ended questions requesting an explanation of the procedure the participant followed to complete a given task.*

Results

To evaluate performance on each display, root mean squared error (RMSE) was calculated from the differences between optimal performance and observed performance. For altitude and heading, the optimal flight path was calculated and RMSE was calculated for each pilot. For power, RMSE was calculated by comparing actual performance against optimal baseline performance (that is, the optimal power settings for straight and level flight).

A repeated measures, one-way ANOVA was used to analyze performance differences between the two displays. For this report, the results will focus on the performance differences for power and altitude between display conditions.

Straight and Level Flight. Analysis showed novices in the Oz display condition deviated significantly less from optimal power settings ($F(1, 9) = 33.148$, $p < .001$) than novices trained on the conventional system (Figure 2). This finding is consistent with differences observed between display conditions in experienced pilots (Smith, Boehm-Davis, and Chong, 2004), and is consistent with the effects of direct perception and direct manipulation (Lintern et al., 1999).

Altitude differences also supported performance advantages in the functional display condition (Figure 3), as novices flying the Oz display deviated significantly less from the flight profile ($F(1,9)=26.403$, $p<.001$). Novices using a functional display flew more efficiently and with greater accuracy.

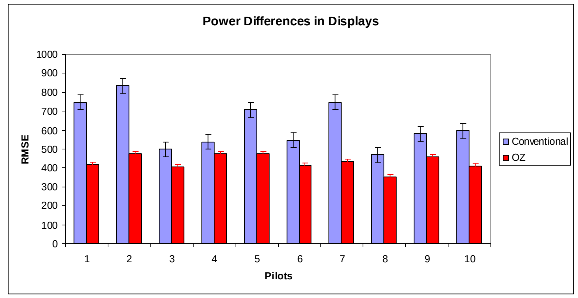


Figure 2. *Power deviations between displays in straight and level flight.*

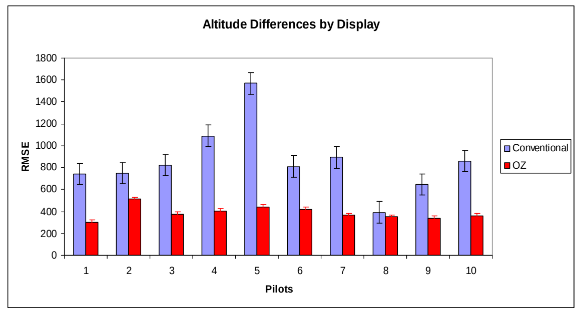


Figure 3. *Altitude deviations between displays in straight and level flight.*

Altitude Ascending

Power use between display conditions was not significantly different. However, novices using the Oz system showed significantly less deviations from the flight profile ($F(1, 9) = 37.465$, $p < .001$).

Standard Rate Turn

Novices performing standard rate turns with a functional display showed significantly less deviations from optimal power settings ($F(1,9)=6.386$, $p<.05$), and from target altitude ($F(1,9)=14.765$, $p<.01$). Novices using a functional display deviated less from optimal power and flew significantly closer to the flight profile. These performance differences may result from the presence of a visual cue indicating the power necessary to maintain straight and level flight. In a prior study, the visual referent seemed to provide a baseline from which pilots could judge power settings. It is possible that novices using the functional display could perceptually judge appropriate power use off of a visual cue of distance from power for straight and level.

Increasing Speed

Novices trained on the Oz display deviated significantly less from optimal power settings (F (1, 9) =4.808, P<.05) when increasing speed (Figure 4) than novices using the conventional display. Novices using the Oz display also deviated less from the flight profile (F (1, 9) =14.688, p<.01) (Figure 5). No significant differences in speed control were observed between display conditions.

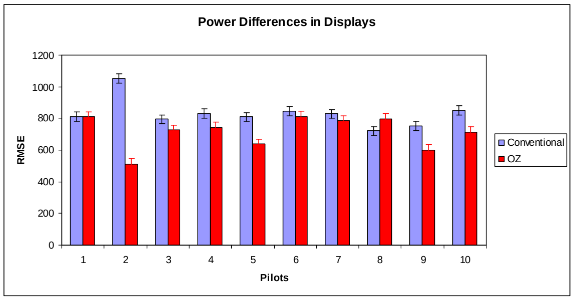


Figure 4. *Power deviations between displays when increasing speed.*

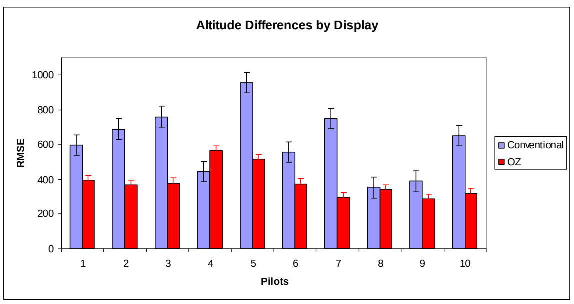


Figure 5. *Altitude deviations between displays when increasing speed.*

Standard Rate Turn while Ascending. Novices trained on the Oz system used significantly more power (F (1, 9) =5.815, p<.05) than their counterparts trained on a conventional display (Figure 6). These maneuvers, however, requires a large amount of the systems available power. It is likely that the amount of power used corresponds with correct operation of the aircraft. Pilots using the functional (Oz) display deviated significantly less (F (1, 9) =23.547, p<.001) from the optimal flight path in altitude (Figure 7) and heading (F (1, 9) =204.26, p<.001) than novices trained on a conventional system. No significant differences in speed control were observed between display conditions, even when novices were given direct instructions to control for speed.

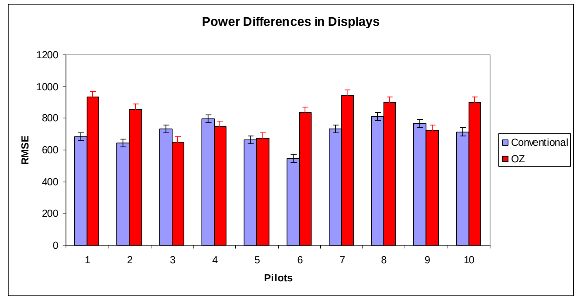


Figure 6. *Power deviations between displays when ascending and performing a standard rate turn.*

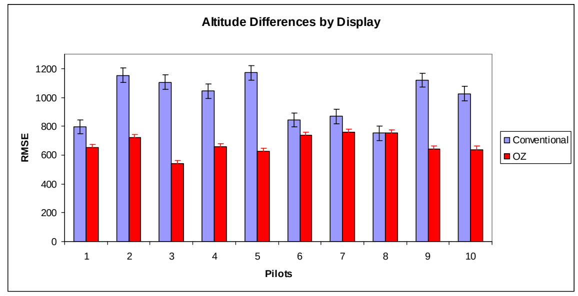


Figure 7. *Altitude deviations between displays when ascending and performing a standard rate turn.*

Discussion

The results offer support for the potential of functional displays to aid in the acquisition of piloting skill. Across all maneuvers, novices using the Oz display were able to maintain a flight profile (altitude and heading) closer to optimal than novices flying conventional displays.

The results point to a potential difference in novice understanding of the functional relationships present during flight. Novices trained on a functional display applied power in patterns consistent with the requirements of the maneuver. This was observed consistently in both straight and level and banking maneuvers, as novices trained on the Oz display were better able to maintain consistent power settings closer to baseline (optimal) conditions. In comparison, novices trained on a conventional display employed varying amounts of power, both within trials and between participants. Lacking a direct functional referent, and given only physical information about required power, novices were required to mentally compute the power necessary to maintain straight and level flight.

Further results also support the use of a functional graphic to communicate essential relationships between system properties. In scenarios requiring the combination of multiple maneuvers (for instance, banking while ascending), novices performing on the Oz system maintained a performance advantage by applying more power with less variability. While novices using the conventional display used less power in relation to baseline power settings (the optimal amount of power for straight and level), they were unable to maintain the required flight profile for the maneuver as accurately as novices employing the functional display. This finding suggests potential differences in understanding the amount of power required to perform a given maneuver between display conditions. The performance differences between displays serve to underscore the potential effectiveness of a functional display. By providing novices with a direct graphical referent to the functional relationship being manipulated, performance in maintaining and controlling that maneuver may be improved.

An interesting approach to these results is to compare the performance of novices in this study with experienced pilots flying the same display. The power settings of novices using the functional (Oz) display in the current study is strikingly similar to that of professional pilots using the Oz display in previous research (Smith, Boehm-Davis, and Chong, 2004). This finding may be construed as novices having a greater understanding of the task requirements, or potentially reflect a greater understanding of the properties of flight. It may also be that the presentation of a direct perceptual graphic in flight provides less skilled pilots a referent to replace mental computations with rule-based, perceptual activity. With either explanation, the results support the notion that displays leveraging direct perception in depicting functional relationships can improve the performance of novice pilots as they execute flight maneuvers.

The challenge for designers of functional displays, then, may be to identify areas within the system where perceptual referents can communicate the functional relationships essential to a novice's understanding of proper system operation. As Rasmussen (1999) noted, an interface designed to support the operator should make performance boundaries visible. Functional graphics designed for areas such as this could help novice users visually perceive a performance envelope that defines the limits of functionally acceptable performance for each user. By leveraging the strengths of the perceptual system, the designer can assist the novice aviator in adapting to complex and abstract relationships present in modern aircraft.

References

Bennett, K. & Flach, J.M. (1992). Graphical Displays: Implications for divided attention, focused attention, and problem solving. *Human Factors, 34(5)*, 513-533.

Dinadis, N., & Vicente, K. J. (1999). Designing functional visualizations for aircraft system status displays. *International Journal of Aviation Psychology, 9*, 241-269.

Lintern, G., Waite, T., and Talleur, D.A. (1999). Functional Interface Design for the Modern Aircraft Cockpit. *The International Journal of Aviation Psychology, 9(3),* 225-240.

Rasmussen, J. (1999). Ecological interface design for reliable human-machine systems. *International Journal of Aviation Psychology, 9,* 203-223.

Rasmussen, J., & Vicente, K. J. (1989). Coping with human errors through system design: Implications for ecological interface design. *International Journal of Man-Machine Studies, 31*, 517-534.

Smith, C.F., Boehm-Davis, D.A., and Chong, R. (2004). The Effect of Functional Information in an Avionics Display. (*In Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting).* New Orleans, LA: Human Factors and Ergonomics Society.

Vicente, K.J. (2002) Ecological Interface Design: Progress and Challenges. *Human Factors, 44(1),* 62-79.

Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics, 22, 4,* 589-606.

Table 1.
Maneuvers Performed by Participants

Presentation Order	**Maneuver**
1	Maintain Straight and Level Flight
2	Ascend/Descend 1000 Ft.
3	Ascend/Descend 1000 Ft.
4	Increase/Decrease Speed from 85-110/110-85
5	Increase/Decrease Speed from 85-110/110-85
6	Bank Left/Right 180 degrees
7	Bank Left/Right 180 degrees
8	Ascend/Descend 1000 Ft., Bank Left/Right 360 Degrees
9	Ascend/Descend 1000 Ft., Bank Left/Right 360 Degrees
10	Ascend/Descend 1000 Ft., Bank Left/Right 360 Degrees, maintain airspeed of 85/105 knots
11	Ascend/Descend 1000 Ft., Bank Left/Right 360 Degrees, maintain airspeed of 85/105 knots

Table 2.
List of Acceptable Feedback Provided by Flight Instructor

Situation	Feedback
Overpowered	"You are overpowered. Reduce power with your throttle."
Underpowered	"You are underpowered. Increase power with your throttle."
Above Altitude	"You are above the required altitude. Lower your altitude."
Below Altitude	"You are below the required altitude. Increase your altitude."
Over Speed	"You are over your target speed. Reduce your airspeed."
Under Speed	"You are under your target speed. Increase your airspeed."
Past Heading	"You are past your required heading. Return to a heading of ___."

COORDINATED CONTINGENCY PLANNING IN THE FACE OF UNCERTAINTY IN THE NATIONAL AVIATION SYSTEM

Philip J. Smith, **Amy Spencer** and **Charlie Billings**
The Ohio State University
Columbus OH

One of the major challenges for strategic planning in aviation concerns uncertainty about weather and traffic constraints, as traffic managers often have to disseminate reroute advisories 2 hours before an expected constraint impacts an airport, and dispatchers file flight plans 60-75 minutes before a flight's departure. When the predictions used to for these plans are wrong, significant inefficiencies (unused airspace and runway capacity from a traffic manager's perspective and delayed flights from a dispatcher's perspective) often result. To make operations more adaptive, new procedures have been developed. These procedures involve using predefined Coded Departure Routes, and are now being extended to include the dissemination of strategic plans that explicitly deal with uncertainty. Through this process, the decision about what departure route to actually use for a flight can be delayed until it is ready to depart, avoiding the need to make an early (and potentially poor) commitment to a departure route that may be unavailable at the time the flight taxis out for departure, while still keeping the dispatcher in the loop.

Background

In order to deal with cognitive complexity, the operation of the National Airspace System (NAS) is distributed among many organizations and individuals. The architecture for this distributed work system can be characterized in terms of the allocation of control and responsibility, and also in terms of the distribution of data, knowledge, processing capacities, goals and priorities. Within this distributed system, one of the most significant challenges is how to coordinate and adapt plans in the face of uncertainty, given that the level of uncertainty changes over time (Smith, Beatty, Spencer and Billings, 2003).

At present, most procedures to use traffic flow management in order to improve coordination must oversimplify consideration of this time-varying uncertainty. This is done by making predictions about the most likely scenario and developing a resultant single plan. Figure 1 is an example of an advisory describing such a plan.

In this paper, we explore enhanced communications between traffic management and the NAS users which allow them to deal more effectively with uncertainty in weather and traffic constraints. Instead of a process that communicates a single plan, a process that is currently being implemented by the FAA traffic managers and dispatchers will make it possible for both traffic managers and dispatchers to communicate constraints and contingency plans. By communicating within this more expressive framework, data and knowledge are shared in an efficient manner at an appropriate level of abstraction, in order to allow both traffic managers and dispatchers to plan the actions under their control in a more informed and realistic manner.

New Solution

Coded Departure Routes (CDRs) are a set of predefined alternative routes for flying between particular city pairs. They were developed by ATCSCC and ARTCC staff in cooperation with the NAS users under the auspices of the FAA's Collaborative Decision Making Program (Beatty and Smith, 2000; Smith, et al., 2001; Smith, Beatty, Campbell, et al., 2003).

These prespecified routes were developed for two reasons. First, there is an 8 letter abbreviation associated with each CDR, making computer entry and communication of that route much faster for FAA and dispatch staff (thus reducing workload and expediting route changes). Second, these CDRs were designed to support a collaborative process for selecting an alternative departure route for a flight when the user preferred route is not available due to a weather or traffic constraint.

More specifically, the initiative that led to the development of CDRs had several underlying motivations. The first was to increase efficiency in communicating changes in the departure route for a flight, speeding up such communication and reducing the associated workload. The second was to develop a collaborative process that was intended to:

- Provide Airline Operations Centers (AOCs) and other NAS users, along with traffic managers at ATCSCC, ARTCCs, TRACONs and Towers with a process for working collaboratively to develop earlier plans for dealing with predicted constraints in the NAS.
- Provide a set of pre-specified alternate departure routes for specific city pairs that had been approved by all of the involved Centers in terms of the impact

on typical traffic flows and constraints.
- Give traffic managers greater flexibility in responding to the often rapidly changing picture regarding available airspace during weather and traffic events, so that departure delays could be reduced.
- Keep dispatchers in the loop through the early identification of the alternate departure routes that might be selected at the time of departure from an airport.

As an illustration, see Figure 2, which shows a scenario involving uncertainty about when a storm cell will close off departures out of DTW via CAVVS, making it desirable to have the CDR from DTW via WINGS available as an alternative departure route. Figure 3 shows an analogous situation for departures out of New York, with weather potentially impacting departures via ELIOT, with COATE as an alternative departure fix. As indicated in the table in Figure 3, the flight was filed by the dispatcher for departure via ELIOT at 1734Z, but was re-cleared for departure via COATE by a traffic manager at New York Center at 1856Z (Smith, et al., 2005). This reroute allowed the flight to depart on time instead of having to wait for the weather to clear.

In terms of making communications more effective, and in reducing coordination time among FAA facilities, CDRs have been quite successful (Smith, 2003). However, the desired improvement in coordination and preplanning between traffic managers and AOCs and other NAS users has not been as effective. As one traffic manager indicated (for his Center's airports):

"The CDRs are usually issued on the taxiway. The pilot then has to contact his dispatcher to see if the flight meets FAR criteria. We have had them taxi back to the ramp to take on more fuel or unload baggage."

While there are Centers and airlines that have developed methods for preplanning when CDRs should be used, this is still the exception and, when it is done, it requires a great deal of effort because communication and coordination is done by phone. Thus, one of the major factors that has limited the effective coordinated use of CDRs has been the lack of software support for communication between traffic managers and the AOCs and other NAS users.

Preplanning For Alternative Departure Routes

To deal with this issue, a number of steps are being taken to improve pre-coordination concerning the use of CDRs for departures from a given airport. Specifically:

- Strategic planning telecons are held every 2 hours, with traffic managers from ATCSCC, ARTCC, TRACON involved, along with air traffic control coordinators representing the NAS users.
- During these telecons, the traffic manager for an ARTCC that anticipates a potential but uncertain constraint (where the uncertainty can be in terms of its timing or location) is asked to provide a prediction about the potential timing and location of the constraint, as well as recommended alternative solutions depending on how the constraint develops. Given the nature of CDRs, such predictions generally focus on potential weather or traffic constraints that are likely to block a given departure direction out of an airport (see Figures 2 and 3).
- The ARTCC traffic manager is also asked to use the FAA's Traffic Situation Display (TSD) to draw a flow constrained area (FCA) indicating the route that may be blocked. For the weather constraint in Figure 2, this area would be drawn directly south of DTW, indicating that the normally preferred routes departing via CAVVS may be blocked by the constraint. For the weather shown in Figure 3, this FCA would be drawn around ELIOT.
- The traffic manager also indicates which alternative routes (CDRs) are expected to be used to expedite departures if and when the constraint does develop.
- This FCA, along with a prediction model for flight trajectories, is then used to identify the flights that are expected to traverse this FCA during the time when that airspace may be constrained.

This information is then included as part of the strategic plan, which is distributed to all FAA facilities and to the NAS users. Specifically, this information includes a graphic indicating the airspace that may be impacted by the constraint, the timeframe during which this could occur, and the recommended alternative routes for which flights should be prepared (if possible). It also contains a list of the flights that are likely to be affected. Below is an example of the information contained in such a strategic plan regarding preparation for an alternative route.

"For flights departing ZNY and ZBW 1600-2200Z, file on J36/J95/J60 if desired, but prepare for possible use of CDRs on J64 and J80 (see FCA004 for flight list)"

Assuming this strategic planning information is received by the dispatcher before preparing the flight plan for a flight (typically 60-75 minutes before departure), the dispatcher must decide whether it is safe to file a route that assumes the constraint will not impact the flight. (In the scenario illustrated in Figure 2, in such a case the dispatcher would file a departure via CAVVS; in Figure

3 the dispatcher would file a departure via ELIOT.) The fact that the strategic plan has an attached flight list further means that only the dispatcher with an affected flight needs to review this advisory.

Given the strategic planning information, the dispatcher would proceed to evaluate that flight for departure using a CDR via WINGS for the scenario in Figure 2 or via COATE for the scenario in Figure 3. If the dispatcher determined that such an alternative route was safe and effective for the flight should the weather impact CAVVS (Figure 2) or COATE (Figure 3) at departure time, then the flight could be pre-approved for and fueled for this alternative route. This information would then be included on the flight release, letting the flight crew know that they could accept a clearance on the filed (user preferred route) or the pre-approved alternative.

Just prior to departure, a traffic manager would then evaluate the situation, leaving the flight on the user preferred route if that was available for a timely departure, or moving it to the alternative CDR if that expedited its departure. This information would then be sent to the airport Tower controller, who would give the flight a clearance for departure on the originally filed route or the alternative CDR, depending upon what the traffic manager had decided.

Note that, in some cases, the dispatcher might choose to not approve the alternative route for some safety or business reason, in which case the flight would either have to take a delay on the ground or the dispatcher would have to request an exception for some other alternative route from traffic management.

Summary

One of the major challenges faced by traffic managers and dispatchers is dealing with uncertainty regarding weather and traffic constraints. To improve performance in the face of such uncertainty, they have begun to develop a system that allows much more adaptive and agile responses as specific scenarios unfold.

The introduction of CDRs represented one important step in this direction, reducing coordination time among traffic managers and reducing communication times among traffic managers, dispatchers, pilots and controllers. This paper describes the next step in trying to make the system even more adaptive, while ensuring that all of the critical parties remain in the loop. This next step involves the creation and dissemination of strategic plans that identify contingencies for dealing with uncertainty.

Under this new procedure, traffic managers share their knowledge by suggesting potential contingencies. Dispatchers input their expertise by determining whether or not to pre-approve these contingencies. Through this process, the decision about what departure route to actually use for a flight can be delayed until it is ready to depart, thus avoiding the need to make an early (and potentially poor) commitment to a departure route that may be unavailable at the time the flight is ready to depart. This makes it possible to clear the flight on a route that expedites its departure, while still ensuring that the dispatcher has been involved in evaluating the safety and efficiency of the final route.

Acknowledgements

This work was supported by the FAA Collaborative Decision Making Program and the FAA Human Factors Research and Engineering Division.

References

Beatty, R. and Smith, P.J. (2000). *Design Recommendations for an Integrated Approach to the Development Dissemination and Use of Reroute Advisories*. Institute for Ergonomics Technical Report #2000-13. Ohio State University, Columbus OH.

Smith, P.J. (2003). *Benefits from the Use of Coded Departure Routes.* Institute for Ergonomics Technical Report #2003-15. Ohio State University, Columbus OH.

Smith, P.J., Beatty, R., Campbell, K., Murphy, M., Jezerinac, J. and Behbehani , T. (2003). Design and evaluation of tools for creation, dissemination of and response to reroute advisories to enhance coordination in traffic flow management. *Proceedings of the 2003 Aviation Psychology Symposium*, Dayton, OH.

Smith, P.J., Beatty, R., Spencer, A. and Billings, C. (2003). Dealing with the challenges of distributed planning in a stochastic environment: Coordinated contingency planning. *Proceedings of the 2003 Annual Conference on Digital Avionics Systems*, Chicago, IL.

Smith, P.J., Klopfenstein, M., Jezerinac, J. and Spencer, A. (2005). Distributed work in the National Airspace System: Providing feedback loops using the Post-Operations Evaluation Tool (POET). In B. Kirwan, M. Rodgers and D. Schaefer (eds.), *Human Factors Impacts in Air Traffic Management.* Ashgate.

Smith, P.J., McCoy, E. and Billings, C. (2001). *Issues in the Use of Coded Departure Routes*. Institute for Ergonomics Technical Report #2001-12. Ohio State University, Columbus OH.

Date: 12/23/2004 12:12   Title: ROUTE RQD /FL   NAME: SNOWBIRD_7
CONSTRAINED AREA: ZDC   REASON: VOLUME

INCLUDE TRAFFIC: ATL/CLT DEPARTURES TO BDL/BED/BOS/HPN/PVD
FACILITIES INCLUDED: ZJX/ZTL/ZDC/ZNY/ZBW
FLIGHT STATUS: ALL_FLIGHTS
VALID: ETD 231208 TO 231630
PROBABILITY OF EXTENSION: MODERATE

REMARKS: AIRCRAFT FILED VIA A761 OR THE ATLANTIC ROUTES ARE EXEMPT
ASSOCIATED RESTRICTIONS: AS COORDINATED.
MODIFICATIONS: ATL/CLT DEPARTURES ONLY.
ROUTES:

ORIG	DEST	ROUTE
ATL	BOS	SPA J14 PXT J191 RBV J222 JFK ORW3
ATL	PVD	SPA J14 PXT J191 RBV J62 J150 HTO JORDN MINNK
CLT	BOS	RDU J55 HPW J191 RBV J222

Figure 1. *Sample reroute advisory assigning specific reroutes instead of preparing for alternative contingencies.*

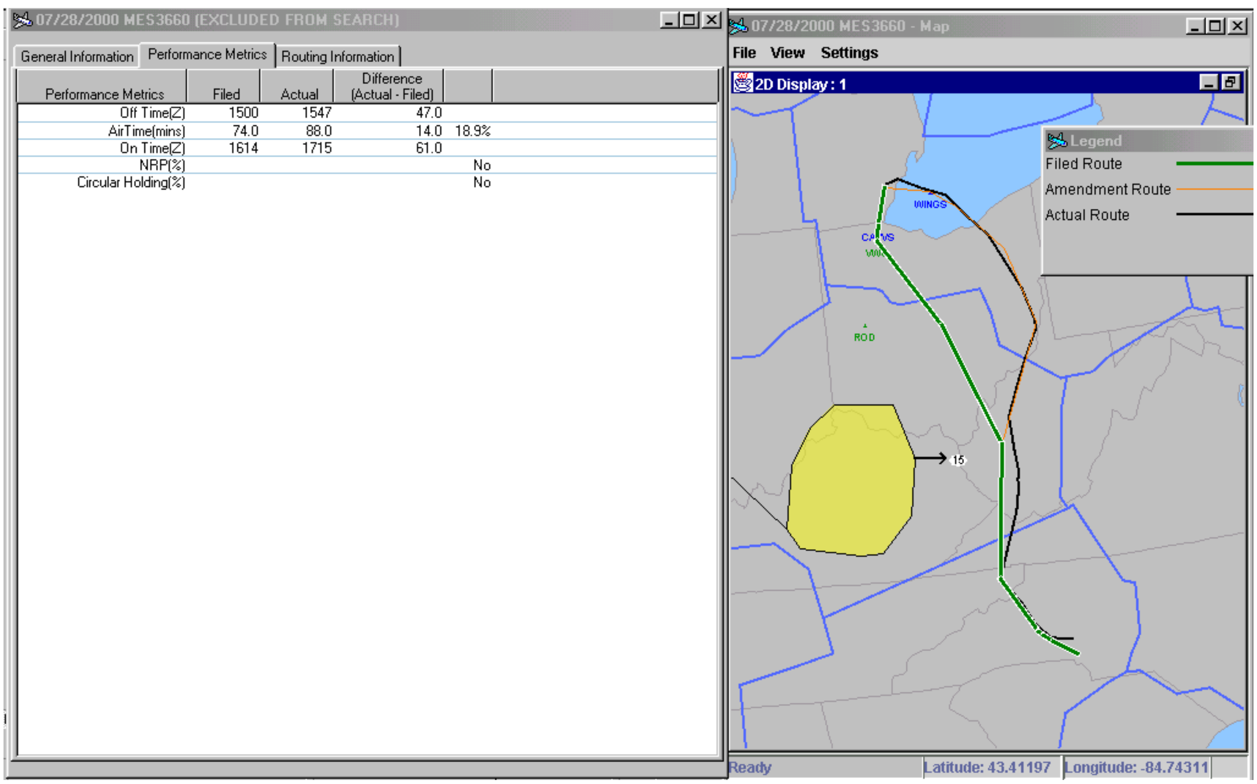


Figure 2. *Initial information on a specific flight.*

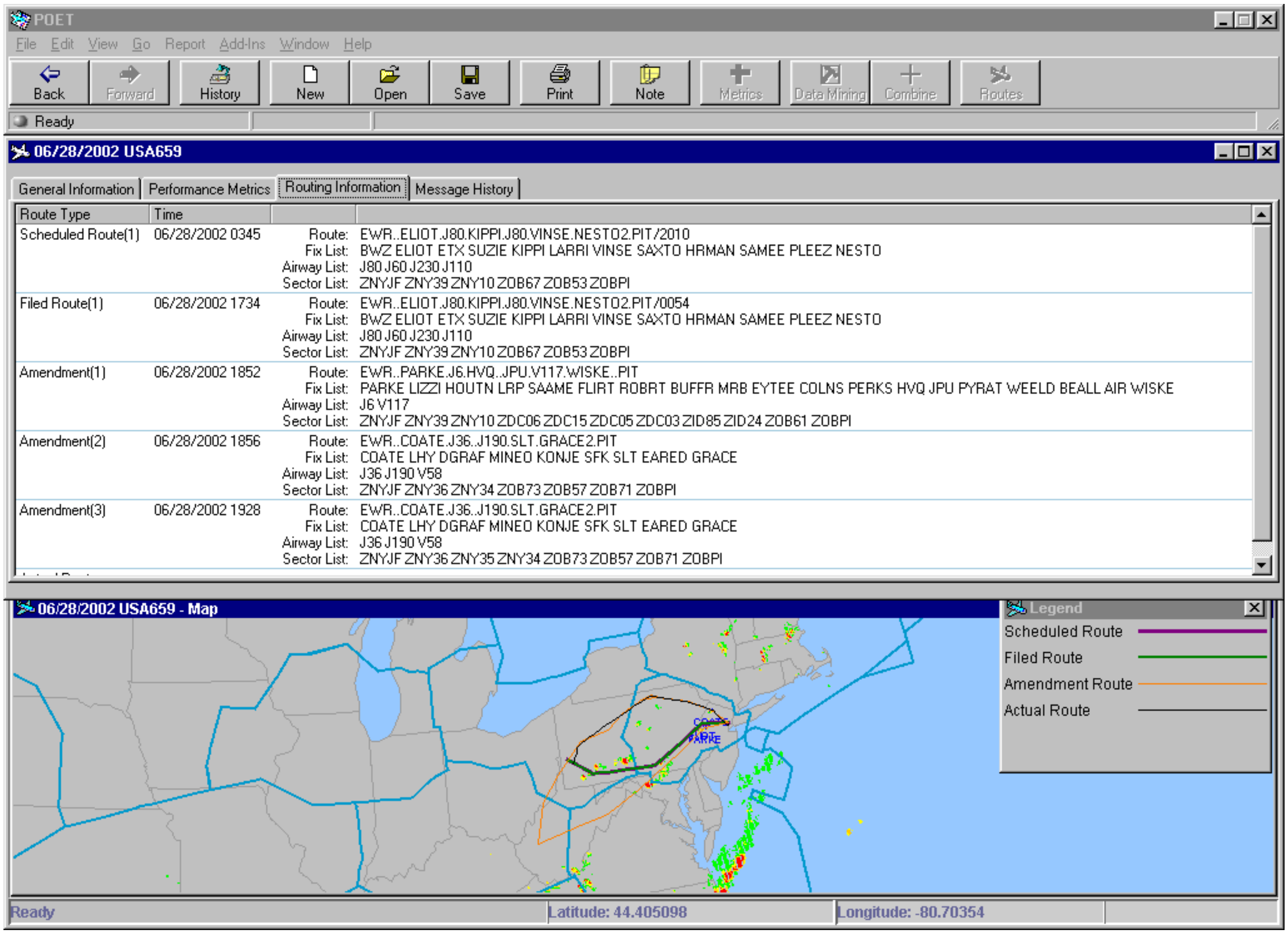


Figure 3. *Flexible routing to expedite departure of a flight from EWR-PIT*

A MIND-REFERENCE FRAMEWORK FOR DESIGN AND EVALUATION OF INTUITIVE AND NATURAL INTERFACES

Iya Solodilova
Dept of Computer Science
University of Bath
Claverton Down
Bath BA2 7AY, UK
I.Solodilova@bath.ac.uk

Gavan Lintern
General Dynamics
5200 Springfield Pike
Dayton
Ohio 45431-1289
glintern@earthlink.net

Peter Johnson
Dept of Computer Science
University of Bath
Claverton Down
Bath BA2 7AY, UK
P.Johnson@bath.ac.uk

The Mind-Reference framework is proposed to address new and existing interfaces at semantic, perceptual and contextual levels. This framework allows us it is possible to distinguish information structures at behavioral, physical and environmental levels. The framework deals not only with how the information is presented on a perceptual level but also, by accounting for a variety of task contexts, how a pilot can interpret that information. A design approach that follows this framework's step-principles produces intuitive and natural interfaces for pilots and offers a benchmark for evaluation of existing interfaces.

The Problem

The need for intuitive and natural interfaces is a primary topic within the debate about the complexity of modern flight interfaces. Here we explore design principles that could be used to develop an intuitive and natural interface and how could such an interface could be evaluated to determine that the presentation of essential information is intuitive to pilots?

Background

The framework detailed here was developed, in part, from a simulator study in which the pilots wore head-mounted video cameras throughout the flight. A modified version of a cued-recall debrief technique (Omodei, Wearing & McLennan, 1997) was applied to conduct pilot interviews using captured video footage. A structured interview during debrief uncovered the cognitive information strategies used by pilots. These methods revealed what is natural and intuitive to pilots as they use everyday information; how they collect, collate and understand information (Solodilova & Johnson, 2005). The framework incorporates principles that, if followed systematically within design and evaluation of a cockpit interface, will lead to an intuitive and natural presentation of information to pilots.

Birth of Framework

The framework consists of a Mind Reference information matrix that specifies structures, strategies, rules and step-principles to follow when designing or evaluating an interface (Figure 1). All dimensions of the framework, identified during the former study (Solodilova & Johnson 2005), are based on the analysis of how pilots work with information from their point-of-view throughout the flight.

The matrix is based on specific elements of information that pilots manipulate to make sense of their 'information space'. These have been termed as *Mind References*, because pilots mentally collate and then store these pieces of information in the mind until they are needed. They are reliable and unchangeable pieces of information that are aligned relative to other, already established pieces of information (Figure 2).

The established pieces of information align into existing information *structures* that are constantly used in the aviation domain, for example the structure of flight stages. *Structures* are aligned Mind References that establish meaningful relationships in information among vast amounts of it.

Strategies are pilots' approaches to and inventive ways of using the information layout to their advantage. Strategies help pilots deal with information effectively, for example to recover from a loss of or a rapid change of information.

Lastly, *rules* are essential guidelines that pilots learn by rote. These are taught to pilots in training and are reinforced through operational practice. Rules guide pilots to information that supports efficient and successful aircraft operation, for example "always be ahead of the aircraft's action".

An Information Matrix

The *Mind Reference information matrix* serves two purposes. Firstly, it shows the information levels at which pilots have problems, thus helping in the evaluation of interfaces to identify potential information problem areas. Secondly, during

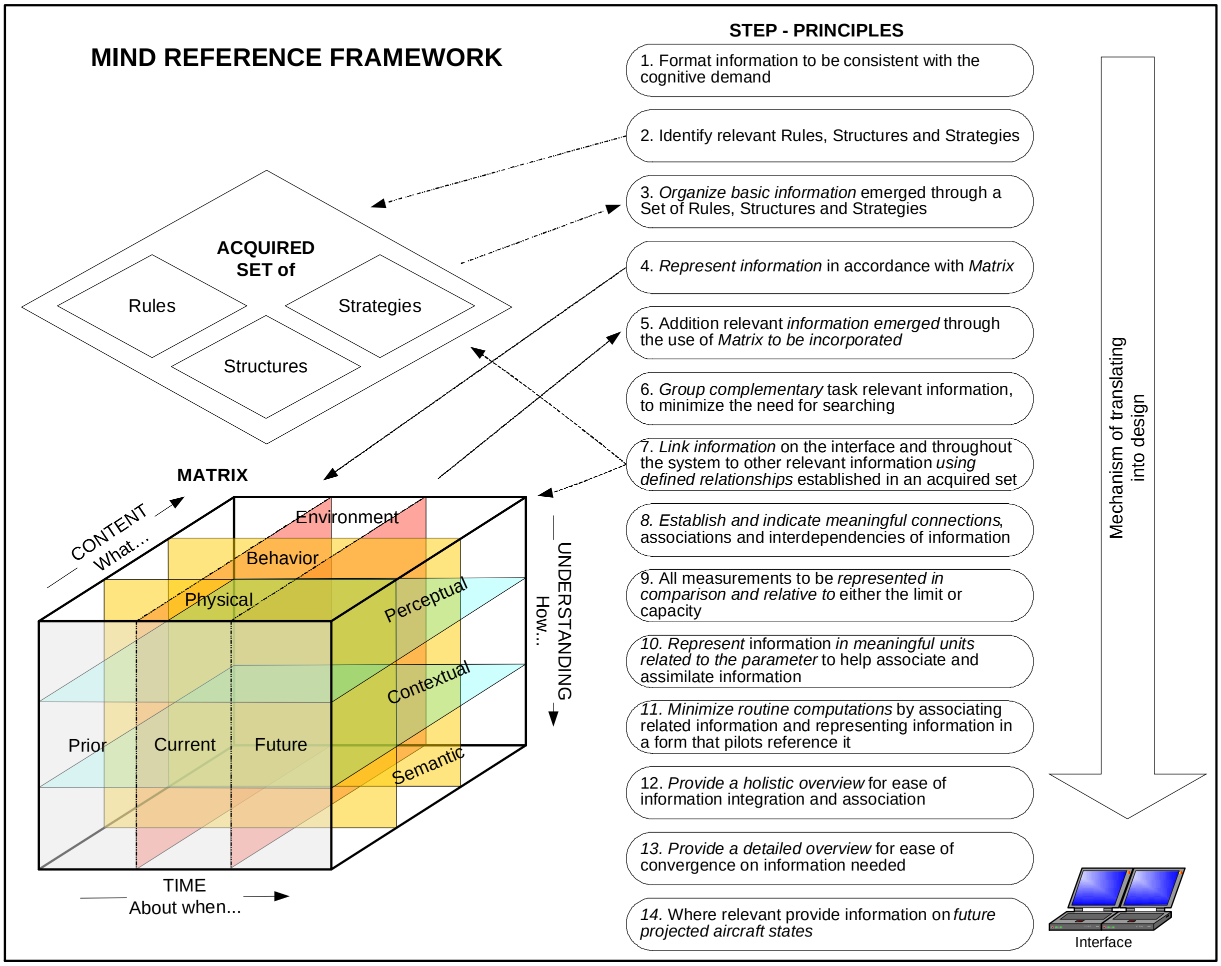


Figure 1. *Mind Reference Framework.*

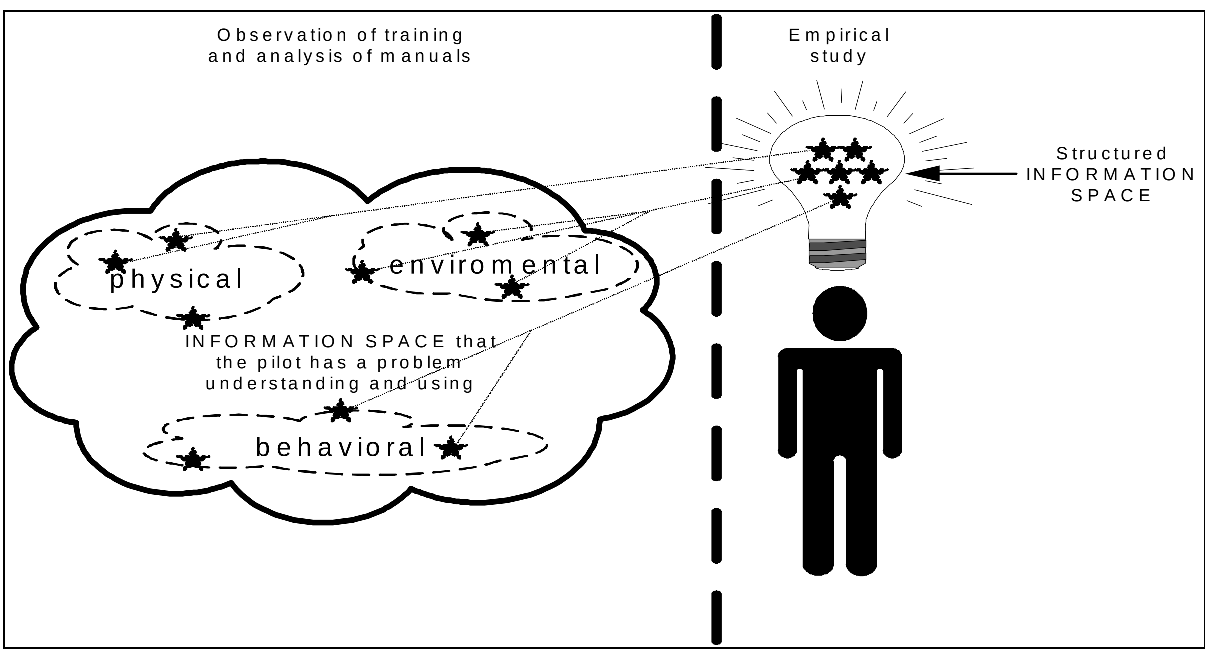


Figure 2.*Information Space.*

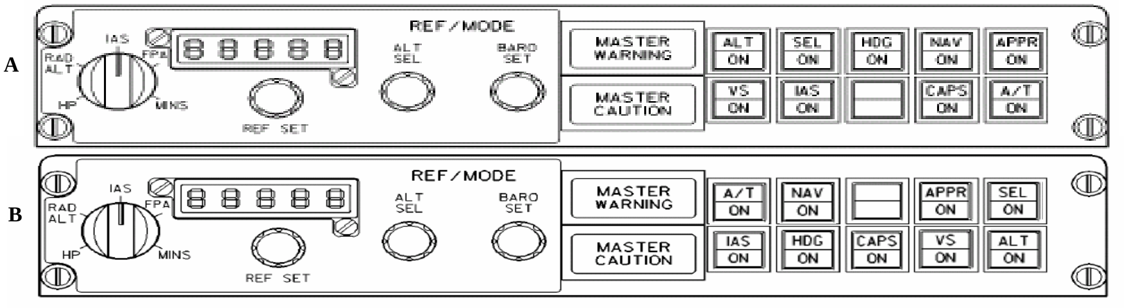


Figure 3. *(A) Original and (B) modified Reference set/mode select panel.*

interface design and evaluation, it directs attention to possible solutions for issues related to information presentation and structure.

The matrix (Figure 1) has three dimensions of information: types of information *understanding*, types of information *content* and *time* dependent information. The dimension of understanding (i.e., how) is composed of three levels at which pilots have problems understanding information: perceptual, contextual and semantic. The information content dimension (i.e., what) consists of three areas of information that pilots need throughout the flight: physical, behavioral and environmental. The third dimension (i.e., when) consists of time dependent information about past, present and anticipated future. These are the three dimensions in which pilots manipulated information.

Use of the Matrix for Evaluation

Several current interfaces from the Hercules aircraft have been chosen as examples to show step-by-step how to assess whether information presentations are unnatural and non-intuitive for pilots. These examples reveal why some existing information presentation solutions are problematic.

Structures

As an example, a reference set/mode select panel (see Figure 3A) is evaluated using the framework. The panel is located on the glare shield in front of each pilot. Starting with the information matrix, it is necessary to first examine the information content using the content dimension (i.e. What) in relation to the set of buttons located on the right side of the panel.

Out of the nine buttons, four select basic behavioral parameters to maintain, such as IAS (Indicated Airspeed), HDG (Heading), VS (Vertical Speed) and ALT (Altitude). The remaining five buttons select more complex automation behavior. For example, the APPR button engages an automation mode to track the selected Instrument Landing System. The SEL button commands the automation to capture selected altitude in climb or descend. The NAV button arms selected navigation mode, and the A/T button engages autothrottle. CAPS is a nonfunctioning button. Thus, in our proposed design, we arranged these buttons into two sets of behavioral instructions, basic and more complex.

The next matrix dimension assesses levels of understanding of information (i.e., How), starting with the perceptual level. This dimension helps to determine the suitability of the button structure. The current structure has no recognizable information structure. Unnecessary introduction of any new structures can create an additional cognitive demand on pilots. The framework helps to identify an information structure that has the same or similar content already ingrained in pilots' minds. The underlying assumption of this approach is that pilots will more easily associate with any new button structure if it conforms to an intuitive, already learned mental structure.

There is already an information structure that reveals aircraft behaviour to the pilot. It is presented on the six standard flight instruments on the panel of most aircraft. These standard instruments are the Airspeed, Turn Co-coordinator, Attitude, Heading and Vertical Speed Indicators, and Altimeter. The Hercules aircraft has these instruments arranged in a specific order on a Primary Flight Display and for our proposed design, we placed these in a single row: IAS, HDG, VS, ALT, as indicators of basic flight response.

Moreover, more complex automation behaviour is already announced at the top of the same display as automation modes, in the following order: A/T, NAV, SEL, APPR (see Figure 5). The only difference is that Autothrottle mode is announced inconsistently. On the Primary Flight Display it is announced as AT, but on the panel as A/T. This annunciation inconsistency needs to be corrected, unless there is a justification for the difference in annunciation.

According to evaluation through the behavioral and perceptual dimensions of the framework, the structure of lines on the reference set/mode select panel would benefit by reflecting the structure on the Primary Flight Display (see Figure 4). Line two should reflect the structure of basic behavior (i.e. IAS, HDG, VS, ALT), as present on the display. The top line should select complex automation behavior modes (i.e. A/T, NAV, APPR, SEL) (see Figure 5) with one additional swap between SEL and APPR buttons. This is dictated by the most basic structure of instruments on the display. Both lines should be ordered and positioned according to the existing structure on the display. For example, the Autothrottle button should be above the IAS button and the SEL button should be above ALT button, because bottom raw buttons (IAS and ALT) select corresponding complex automation behavioral modes (see Figure 3B).

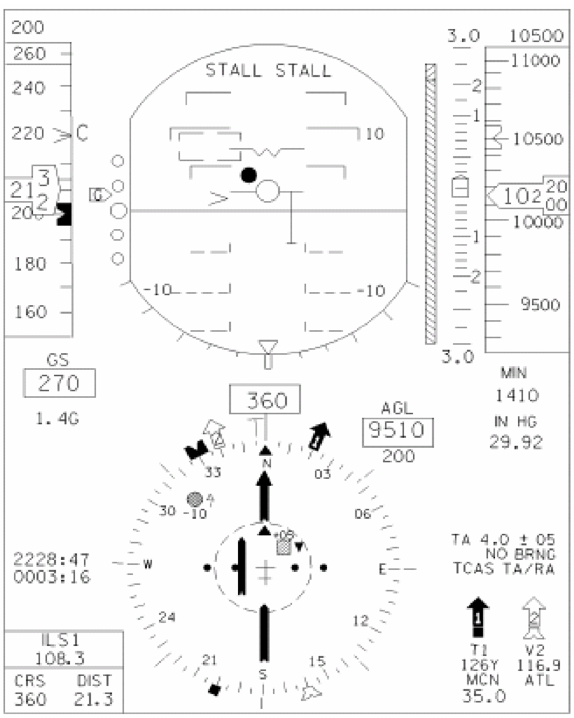


Figure 4. *Primary Flight Display*

Figure 5. *Top of the Primary Flight Display.*

We have placed the CAPS button in the middle of the bottom row because that is the position of this symbol on the Primary Flight Display (see Figure 5 – vertical line on the middle of the display crossing the horizon line). However, the suitability of its position on the Primary Flight Display will be discussed in the next section, the semantic dimension of the matrix.

Based on these two dimensions of the Matrix (What and How) out of three available, it can be established that the right hand side of a reference set/mode select panel does not follow established information structures and can be improved based on an existing structure of the same information that is familiar and in constant use by pilots.

Over a century of operation, aviation has established *structures*, to which pilots constantly refer. Among those are flight stage sequence, Air Traffic Control call order and other established configurations, such a T-instrument layout. These types of information structures are natural and familiar to all pilots and should be used in design, unless more cognitively efficient solutions can be discovered.

Semantic level: consistency in application

There are problems in the modern cockpit that are hard to identify with the evaluation methods currently used in industry (Singer, 2001; Newman, & Greeley 2001). Pilots have difficulty understanding and interpreting available information (Sarter & Woods 1994; 1995). The semantic level of our matrix offers a solution. Although most of the information on the Primary Flight Display has a perceptually plausible interpretation, some features in close proximity to each other offer contradicting meanings.

Consider the following: A *'Fly towards'* or *'Fly-to'* principle has been introduced to the modern cockpit. Most features on the Primary Flight Display, such as Flight Director cues or TCAS RA (Traffic Collision Avoidance System) comply with this principle. However, when features are presented side by side and do not follow the same principle (i.e. How – semantic level), confusion can result at a critical moment of operation. The semantic level of the Matrix offers evaluation of such presentation and helps to bring interpretation of the display into one 'semantic principle'.

Several features on the Primary Flight Display comply with '*Fly*-to' principle, e.g., Glideslope Indicator, CAPS speed bug and Integrated Flight Director. The Integrated Flight Director, for example, gives the pilot precise trajectories for ease of flight control. However, other features on the same display do not follow the same principle and in fact demand the opposite response (i.e. 'fly away') from the pilot. This can create confusion and an incorrect response on the part of the pilot. The features that do not follow the '*Fly*-to' principle are the Speed Error Tape, the Acceleration Cue and the CAPS Distance tape. If the Speed Error Tape is below the Climb/Dive marker it means the aircraft has deviated from and is below the required speed. The pilots' response should be to increase speed. However, if the pilot interprets this is as a '*Fly To*' principle, which is possible since the feature is attached to another feature that complies with this principle, the pilot could potentially respond incorrectly and put the aircraft in an undesired position.

There are also less obvious problems that are semantic in nature. These would not appear to be problems if the pilot were to memorize the meaning behind each feature or word. However, if the pilot forgets the feature's meaning and needs to search for a possible logic behind each feature to establish what it means, errors are likely. Consider the following example.

A Non-Directional Beacon is represented as a triangle. Although the Non-Directional Beacon does not provide direction, the triangular shape of its symbol could be interpreted as a directional arrow. In contrast, a Directional Beacon is represented as a circle, which does not suggest direction via its visual properties. A better solution would be to exchange these symbols. The Non Directional Beacon could be represented as a circle to indicate the 'point of origin' for a signal and the Directional Beacon could be represented as a triangle so that the directional cue was embedded as a visual property.

The semantic level of the matrix directs the evaluation team to identify whether the symbology and presentation of information is optimal, familiar to pilots and has no double meaning behind it.

There is a similar semantic problem related to interpretation of signs on the Head Up Display. However, here the third contextual level of the matrix's understanding dimension directs attention to interpretation of symbology that can be influenced by the context in which it is presented.

The Pitch recovery feature, termed the 'Chevron pairs' (^^) indicates that the nose of the aircraft is high. In doing so, it clashes with the '*Fly-to*' principle that is also applied on this display. Furthermore, the pilot can interpret this feature as a command to 'recover up', because it appears as two arrows pointing upwards. In following this signal, the pilot would put the aircraft in an unusual attitude.

The more problematic issue with 'Chevron pairs' is that there is a similar feature that indicates a nose-low attitude, but is represented as a single Chevron (^) and actually this time does show the recovery direction (i.e. 'Fly-to' principle). If the pilot misinterprets one of these chevrons (Figure 6), the aircraft would be recovered in the wrong direction which again would result in the unusual attitude.

The semantic level of the matrix emphasizes the importance of avoiding symbols that have double meanings or that, due to context, can be interpreted in different or contradictory ways.

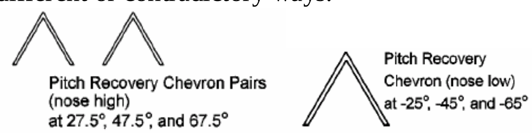


Figure 6. *Chevrons*

The use of Framework step-principles

The framework's step-principles guide design and evaluation of an interface in a step-by-step fashion. The detailed application of the principles in design of a new interface is laid out in previous work (Solodilova, Lintern & Johnston, 2003; Solodilova & Johnson, 2004).

During evaluation, the interface should be judged against each step-principle. Here, we provide an example of how a designer would apply the step-principles for interface evaluation.

Environment level: information proximity

The framework's principles 6, 7 and 8 emphasize the importance of linking and grouping complementary information as well as representing meaningful relationships between related information. The location of interdependent information that is spatially separated and without other forms of association should be identified during evaluation, especially if this information is naturally and routinely used in conjunction with each other.

The readings of barometric pressure and altitude are interdependent pieces of information. On the Head-Up Display, barometric pressure is separated from altitude even though the accuracy of the altitude reading depends on barometric pressure. The seriousness of this problem has noted in a survey of forty-six pilots, where nearly half of the pilots reported that they had set the wrong barometric pressure or had seen another pilot do so (Demagalski, et al 2002).

The DME (Distance Measuring Equipment) information is similarly away from the other related navigational data, such as the source of navigation information. If the pilot reads the navigation information correctly, but it is from the wrong source, that information is of no use.

Both of the problems described above were identified via the framework's evaluation step-principles.

Step-principle 9: 'relative to'

Step-principle nine of the framework proposes that all measurement related information has to be *represented in comparison to and relative to* either the limit or capacity of the parameter it represents. The automation has operational boundaries that are programmed into the system, some of which pilots need to know. During climb, for example, the selected NAV (Navigation) or ALT (Altitude) automation mode may not capture course or altitude respectively if there is a large deviation. The automation tolerates the deviation only within specific limits. The altitude will only be captured within 10% of the rate of climb and the course will be captured only within 5% of the selected course but not otherwise. These limits are not announced to the pilot who can remain unaware of why the automation did not accomplish the commanded operation (i.e., capture NAV or ALT modes).

The above example illustrates the application of step-principle nine and the importance of presenting limits and operational tolerances for automation. Those limits should be identified during design. If not identified during design, they should be detected during the evaluation.

Conclusion

The framework outlined here was developed initially from systematic observation and analysis of operational video data of pilots during simulated flights. The analysis emphasized the use of information from the pilot point-of-view. This emerging framework offers guidance for both design and evaluation of information structures behind modern cockpit displays.

Continuing advances in flight displays and automation have imposed new ways to fly and new ways to interpret information on pilots, but further innovation of the information structures behind the displays is not always desirable. Instead, there is considerable advantage in returning to the basic concepts of flight and the basic strategies of piloting to understand the mental processes that have become ingrained within the aviation profession. New technology and automation offer radically new ways of representing information and of controlling an aircraft but the design of these technologically advanced systems must be constrained by mental structures that pilots find natural.

The modern cockpit of the Hercules is not the only one that can benefit from use of the Mind Reference framework for design and evaluation. Modern commercial aircrafts, such as Airbus 320 and Boeing 777, have been evaluated using this framework and similar problem areas were found in cockpit interfaces, where improvements can be made to make interfaces more natural and intuitive to pilots.

Acknowledgments

This work is being funded by the EPSRC (grant number GR/R40739/01) and supported by QinetiQ and Westland Helicopters.
We would like to thank the pilots, who participated in the flight observations study and in particular Squadron Leader Mark Thoresen of 85WG RAAF Base Richmond, NSW, Australia, for organizing and making this research possible.

References

Demagalski, J., Harris, D., Salmon, P., Stanton, N. A., Young, M. S., Marshall, A., Waldman, T., & Dekker, S. W. A. (2002). Design induced errors on the modern flight deck during approach and landing. In: *Proceedings of HCI-Aero 2002*, MIT, Cambridge, MA, pp.173-178.

Newman, R. L. & Greeley, K. W. (2001). *Cockpit Displays: Test and Evaluation*. Aldershot: Ashgate.

Omodei, M., Wearing, A., & McLennan, J. (1997). Head-mounted video recording: A methodology for studying naturalistic decision making. In R. Flin, E. Salas, M. Strub and L. Martin (Eds.) *Decision Making Under Stress: Emerging Themes and Applications*. (pp. 137-146) Aldershot: Ashgate.

Sarter, N. B. & Woods, D. D. (1994). Pilot Interaction with cockpit Automation II: An Experimental Study of Pilots' Model and Awareness of the Flight Management System. *The International Journal of Aviation Psychology*, 4 (1), 1-28.

Sarter, N. B. & D. D. Woods (1995). How in the world did we get into that mode? Mode error and awareness in supervisory control. *Human Factors*, 37 (1), 5-19.

Singer, G. (2001). Minimising pilot error by design: are test pilots doing a good enough job? *Human Factors and Aerospace Safety*, 1(4), 304-321.

Solodilova, I., Lintern G., & Johnston, N. (2003). The Modern Commercial Cockpit As a Multi-Dimensional, Information-Action Workspace. In: *Proceedings of the 12th International Symposium on Aviation Psychology*, 14-17 April 2003, Dayton, Ohio, USA, pp.1096-1101.

Solodilova, I., & Johnson, P. (2004). 'Mind References' in the Glass Cockpit: An Experimental Study. In: *Proceedings of the HCI Aero 2004*, Toulouse, France, [CD-ROM].

Solodilova, I., & Johnson, P. (submitted). Uncovering the Information Needs in Complex Aerospace Systems. *CiD 2005 Complexity in Design and Engineering.*

VALIDATION OF A MODERN AVIATION PSYCHOLOGY TEST BATTERY USING AERTIFICIAL NEURAL NETWORKS: FIRST RESULTS OF TWO PILOT STUDIES

Markus Sommer
Dr. G. Schuhfried GmbH
Mödling, Austria

The present paper deals with the problem of data integration in the context of aviation psychological assessment. In the first pilot study 99 pilot applicants completed a comprehensive test battery. The general judgment of the candidates' performance in a flight simulator served as an external criterion. To examine the predictive validity of this test battery, both a discriminant analysis as well as an artificial neural network were calculated and compared with each other with regard to classification rate, stability, and their respective differentiability of suited and not suited applicants based on their success probabilities. The results of this first pilot study demonstrate that artificial neural networks outperform classical methods of statistical judgment formation with regard to classification rate and differentiability of suited and not suited applicants based on their success probabilities. In the second study 264 applicants for the position of a commissioned officer in the air force completed a smaller test battery, which also included measures of personality traits. The general judgment of the candidates' performance in a flight simulator served as external criterion. To examine the predictive validity of this test battery, a discriminant analysis as well as a logistic regression analysis and artificial neural network were calculated and compared with each other with regard to classification rate, stability, and capacity to separate correct and incorrect classifications. The results of this second study replicate the finding of the first pilot study by demonstrating, that artificial neural networks result into higher classification rates and a better differentiability of suited and not suited applicants based on their success probabilities. Based on these results it is concluded, that artificial neural networks provide a valuable tool for the selection of pilots which increases the objectivity and precision of diagnostical judgments derived from standardized test batteries.

Theoretical Introduction

The main selection criteria for individual tests as well as test batteries used to select pilot applicants are the criterion validity, the overall cost of testing and time requirements. The selection of the respective tests can be based on recommendations of the Joint Aviation Requirements for Crew Licensing 3 (JAR-FCL3) and validation studies. Naturally, the derivation of decisions from a test battery requires a sufficiently high correlation between the tests and the criterion variable. However, recent metaanalysis (cf. Hunter & Burke, 1994; Burke, Hobson & Linsky, 1997) indicates, that the correlation coefficients between a single test and the criterion measure don't exceed an absolute value of .30. There are a variety of causes for this, ranking from a lower reliability of the criterion- or predictor variables (Lienert & Raatz, 1998), an attenutation of the variance in the predictor variables due to selection (Lienert & Raatz, 1998) to the lack of symmetry between the generality of the predictor variables and the generality of the criterion variable. With regard to the later cause Wittmann and Süß (1997), Ajzen (1987) and Ree and Carretta (1996) pointed out, that for more general and global criteria such as successful performance in a flight-simulator or an aviation educational program, aggregate measures such as general ability ("g") are better suited for prediction than more specific predictors. Thus, one way to handle this problem is to combine the available information about an applicant to generate a prediction about her or his success. In general, one can resort to various methods of statistical judgment formation in order to do so. But classical methods of statistical judgment formation, such as the discriminant analysis or the regression analysis, are vulnerable to violations of their statistical assumptions and often lack stability in cross-validation in practical applications (cf. Bortz, 1999; Brown & Wickers, 2000). A promising alternative is the use of artificial neural networks. This statistical method has few requirements with respect to data characteristics and has proven to be a robust procedure for pattern recognition tasks (Bishop, 1995; Kinnebrock, 1992; Mielke, 2001; Rojas, 2000; Warner & Misra, 1996). In a previous study Griffin (1998) evaluated artificial neural networks with regard to their ability to predict naval aviator flight grades in their primary phase of flight training using a test battery which primarily consisted of psychomotor tests. Griffin's results indicated that artificial neural networks resulted in a higher validity coefficient compared to the multiple linear regression analysis. However the difference did not reach statistical significance. In line with the current literature on neural networks (Bishop, 1995), the author attributed this result to the lack of non-linear relations between the chosen predictor variables and the criterion variable. Based on this result the aim of the present study is to compare linear discriminant analysis and a neural network with respect to classification rate and generalizability using a more comprehensive test battery to enhance the possibility of non-linear interactions between predictor variables and the criterion measure.

Study 1

Method

The pilot applicants of an airforce took a comprehensive test battery measuring inductive reasoning (AMT), spatial ability (A3DW), attention (COG), reactive capacity (DT), verbal (VERGED) and visual (VISGED) memory and sensomotor coordination (SMK). A total of eight predictor variables can be derived from this test battery consisting of the main variables of each test. In addition all applicants were subjected to a flight-simulator and separated into more or less successful applicants based on a global rating of their flight-simulator performance.

Sample

The sample encompasses 104 pilot applicants in the course of a pilot training. The complete data of 99 pilot applicants are provided. All the candidates are men between 16 and 25 years of age, with an average age of 20.4 years and a standard deviation of 1.85 years. One of them (1%) had completed just 9 years of school but no vocational training, while 19 candidates (19.2%) had completed a vocational school. 74 candidates altogether (74.7%) provided a high school leaving certificate with university entrance permission, and five candidates (5.1%) graduated from university or college. 53.4% of the sample received a positive global evaluation of their flight simulator performance.

Results

The calculation of the discriminant analysis was carried out with SPSS 10.0. The results indicate, that the discriminant analysis is unable to separate successful and less successful pilot applicants based on their test scores (Wilks-Lambda=.851, df=8, p=.059; Box-M: F=1.363, p=.072). Altogether 69.7% of the total sample were classified correctly. 81.1% of the successful pilot applicants and 56.5% of the not successful pilot applicants were classified in accordance with their global rating obtained in the flight simulator. When the candidates are to be assigned correctly to the two groups, the a priori random rate according to Brown and Tinsley (1983) is situated at 50.96%.

A ”jackknife“ validation was carried out to examine the stability of the results. This is a commonly used method to determine the generalizability and the stability of the results of a discriminant analysis in case a second independent sample is lacking (Brown & Wicker, 2000; Hagemeister, Scholz, & Westhoff, 2002). In this “jackknife” validation 54.55% of the candidates were classified correctly. 56.6% of the candidates whose performance at the flight simulator had been considered suited and 52.2% of those whose performance had been considered unsuited were classified correctly. When the candidates are to be assigned correctly to the two groups, the a priori random rate according to Brown and Tinsley (1983) is situated at 50.17%. Figure 1 shows the distribution of the probability to be judged as successful in the flight-simulator according to the “jackknife” validation of the discriminant analysis.

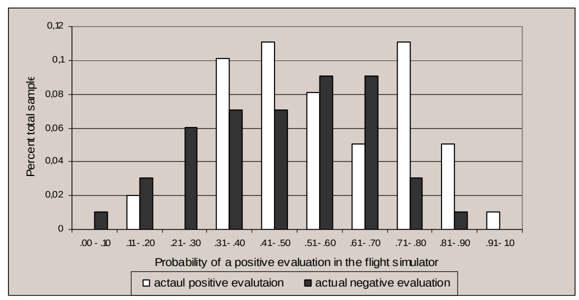


Figure 1. *Distribution of the sample's classification probability according to the "jackknife" validation for the discriminant analysis*

In case the test battery should be used as a screening instrument we would merely take probabilities to receive a positive global evaluation below .25 and over .74 into account. In this case a total of 26.26 % of the pilot applicants can be classified with a reasonably high level of security. The classification rate amounts to 73.08%. As can be seen in figure 1, the majority of incorrect and correct classifications take occur at a rather low level of security.

The calculation of the artificial neural networks was carried out with the program Matlab 6 (Nabney, 2002). The artificial neural network at hand is a multi-layer perceptrone with one hidden layer of five units. The number of "hidden" layer units was determined on the basis of a comparison of various network architectures with respect to parsimony, classification rate and stability. The input layer encompassed eight units representing the individual test scores, while the output layer represents the criterion variable. A feed-forward connection is realized within the neural network. Softmax is used as transformation function. Based on Masters (1995) recommendation, "scaled conjugate gradient" was chosen as training algorithm. The artificial neural network yields a classification rate of 79.80%. 83.02% of the candidates with a positive evaluation of their flight simulator performance and 76.09% of those with negative evaluation were classified correctly. When the candidates are to be assigned correctly to both groups, the a priori random rate according to Brown and Tinsley (1983) is situated at 49.50%.

A "jackknife" validation was realized to examine the stability of the results, which is a commonly used method to determine the generalizability and the stability of the results from an artificial neural network in case a second independent sample is lacking (Bishop, 1995; Michie, Spiegelthaler & Taylor, 1994; Dorffner, 1991). The classification rate according to the "jackknife" validation amounts to 73.74%. 81.13% of the candidates with a positive evaluation of their flight simulator performance and 65.22% of those with a negative evaluation were classified correctly. When the candidates are to be assigned correctly to the two groups, the a priori random rate according to Brown and Tinsley (1983) is situated at 50.67%. Figure 2 shows the distribution of the probability to be judged as successful in the flight-simulator according to the "jackknife" validation of the artificial neural network.

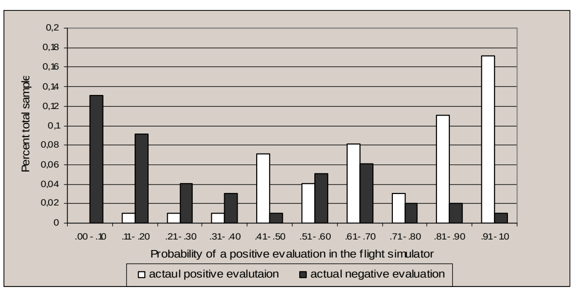


Figure 2. *Distribution of the sample's classification probability according to the "jackknife" validation for the artificial neural network*

In case the test battery should be used as a screening instrument one would merely take probabilities to receive a positive global evaluation below .25 and over .74 into account. In this case a total of 61.61% of the pilot applicants can be classified with a reasonable high level of security. The classification rate amounts to 88.52%. The majority of correct classifications are thus made with high level of certainty, while incorrect classifications were made with a rather low level of certainty. When the candidates are to be assigned correctly to both groups, the a priori random rate according to Brown and Tinsley (1983) is situated at 50.67%.

In case only the most suitable pilot applicants should be selected, cut-off values for the probability to succeed in the flight simulator can be used. Figure 3 shows the percentage of pilot applicants with a positive (doted line) and negative (black line) global evaluation of their flight simulator performance, which reached or exceed a certain cut-off value.

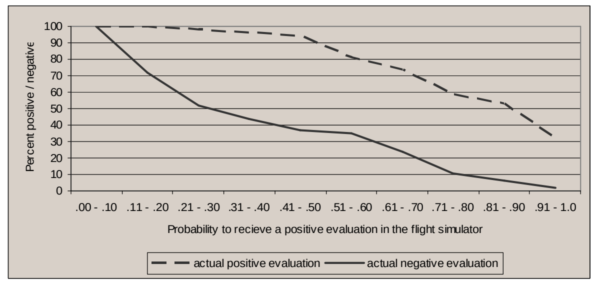


Figure 3. *Percentage suited applicants (doted line) and unsuited applicants (black line) at a given cut-off value*

If a cut-off value of >.70 is chosen, 36.36% of the pilot applicants will be chosen, which includes 58.49% of the pilot applicants, who received a positive global evaluation in the flight simulator and 10.87% of the pilot applicants with a negative global evaluation in the flight simulator. The classification rate amounts to 86.11%. In case a cut-off value of >.80 is applied, one would choose a total of 31.31% of the pilot applicants. Among the chosen applicants there are 52.83% of all pilot applicants, who received a positive global evaluation in the flight simulator and 6.52% of all pilot applicants with a negative global evaluation in the flight simulator. The classification rate amounts to 90.32 %. In practical applications the decision on the cut-off value will be due to the required selection rate as well as the resulting classification rate.

Discussion

The results of this initial pilot study show that artificial neural networks feature an improved classification rate and a better differentiability of correct and incorrect classifications based on the classification probability of the subjects compared to classic methods such as the discriminant analysis. Furthermore, both statistical methods of judgment formation show a comparable high stability of their results. However, the result obtained with the discriminant analysis does not lend itself to a practical application in pilot selection due to the high number of false positive decisions which would result into increased costs for the airforce.

Study 2

Method

All respondents took a test battery measuring figural-inductive reasoning, verbal reasoning and arithmetic computation as well as the personality traits precision level, decisiveness, aspiration level and target discrepancy. Thus a total of seven predictor variables can be derived from this test battery.
In addition all applicants were subjected to a flight-simulator and separated into more or less successful applicants based on a global rating of their flight-simulator performance.

Sample

The sample consisted of 264 male applicants for the position of a commissioned officer in the airforce. 50% of the sample received a positive global evaluation of their flight simulator performance.

Results

The calculations were carried out with SPSS 10.0. A linear discriminant analysis is calculated to predict the applicants' global rating of their flight-simulator performance. The results of the discriminant analysis reveals a violation of the homogeneity-assumption of the variance-covariance matrices which is an essential requirement of the linear discriminant analysis (Box-M: $F=7.214$ $p<.001$). Therefore a logistic regression analysis is used to evaluate the predictive validity of the test battery. Using the method "Enter" the analysis resulted into a -2 Log Likelihood value of 340.127 with $Chi^2=25.855$; $df=7$; $p=.001$. The classification rate amounts to 62.1%. 65.9% of the successful applicants and 58.3% of the less successful applicants are correctly classified.

A "jackknife" validation was performed to examine the stability of the results. In the jackknife validation the classification rate amounts to 58.9%. 62.1% of the successful applicants and 54.5% of the less successful applicants were classified correctly. Figure 4 shows the distribution of the probability to be judged as successful in the flight-simulator according to the "jackknife" validation of the discriminant analysis.

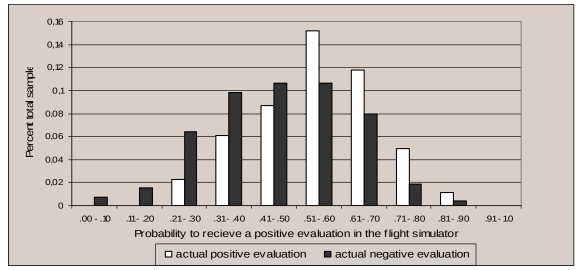


Figure 4. *Distribution of the sample's classification probability according to the "jackknife" validation for the logistic regression analysis*

In case the test battery should be used as a screening instrument one would merely take probabilities to receive a positive global evaluation below .25 and over .74 into account. In this case a total of 3.8 % of the applicants can be classified with a reasonably high level of security. The classification rate amounts to 90%. As can be seen in figure 4 the majority of the classifications are made with a rather low level of security of the classifications.

The calculation of the artificial neural networks was carried out with the program Matlab 6 (Nabney, 2002) using a multi-layer perceptrone using complete feed-forward connections with one hidden layer consisting of five hidden layer units, one input layer with seven units to represent the predictor variables and one output layer with a single unit to represent the criterion measure. The number of "hidden" layer units was determined on the basis of a comparison of various network architectures with respect to parsimony, classification rate and stability. Softmax is chosen as the activation function while quickprop served as training algorithm. The artificial neural network yields a classification rate of 81.7%. 78.6% of the candidates with a positive evaluation of their flight simulator performance and 84.8% of those with a negative evaluation were classified correctly.

A "jackknife" validation was performed to examine the stability of the results. In the validation the classification rate amounts to 75.2%. A total of 78.6% of the candidates with a positive evaluation of their flight simulator performance and 71.8% of those with a negative evaluation were classified correctly. Figure 5 shows the distribution of the probability to be judged as successful in the flight-simulator according to the "jackknife" validation of the artificial neural network.

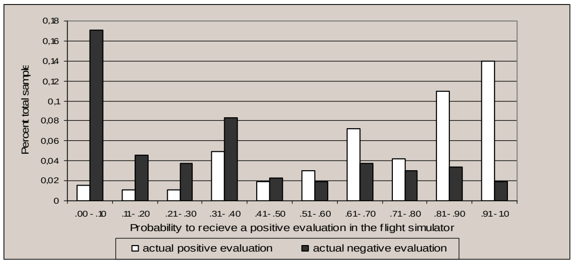


Figure 5. *Distribution of the sample's classification probability according to the "jackknife" validation for the artificial neural network*

For screening purpose one would merely take probabilities to receive a positive global evaluation below .20 and over .70 into account. In this case a total of 54.5% of the applicants can be classified with a reasonable high level of security. The classification rate amounts to 85.42%. As can be seen in figure 5 the majority of correct classifications are thus made with high level of certainty, while incorrect classifications were made with a rather low level of certainty.

If one wants to reduce the amount of unsuited applicants by selecting only the best candidates, cut-off values for the probability to succeed in the flight simulator can be used. Figure 6 shows the percentage of applicants with a positive (doted line) and negative (black line) global evaluation of their flight simulator performance, which reached or exceed a certain cut-off value.

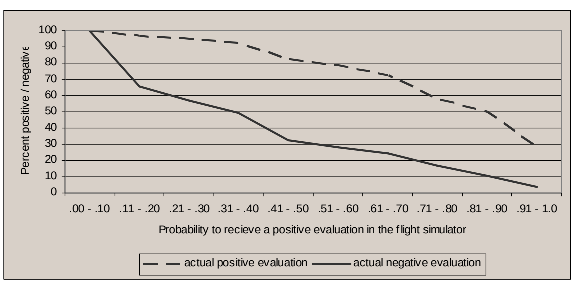


Figure 6. *Percentage suited applicants (doted line) and unsuited applicants (black line) at a given cut-off value*

In case a cut-off value of >.70 is chosen, 37.5% of the applicants will be chosen, which includes 58.3% of the applicants, who received a positive global evaluation in the flight simulator and 16.7% of the applicants with a negative global evaluation in the flight simulator. The classification rate amounts to 77.8%. In case a cut-off value of >.80 is chosen, the selection rate amounts to 30.3%. This includes 50% of the applicants, who received a positive global evaluation in the flight simulator and 10.6% of the applicants with a negative global evaluation in the flight simulator. The classification rate amounts to 82.5 %. Hoewever, in practical applications the cut-off value will have to be chosen based on the desired selection rate and classification rate.

Discussion

In the second study artificial neural networks once more proved to be as stable as classical methods of statistical judgment formation such as the logistic regression analysis. Furthermore, artificial neural networks are even applicable in cases where more traditional methods of statistical judgment formation cannot be applied due to violations of their assumptions. With regard to classification rate and differentiability of correct and incorrect classifications based on the classification probability of the subjects artificial neural networks yielded better results as indicated by the classification rate and the possibility to select the best candidates based on a reasonably high success probability in the criterion measure.

General Discussion

The results obtained in both studies demonstrate, that artificial neural networks outperform classical methods of statistical judgment formation with respect to the magnitude of the classification rate as well as a clearer differentiability of correct and incorrect classifications based on the classification probability of the respondents. Furthermore, artificial neural networks featured a satisfying stability in both studies as indicated by the results obtained in the jackknife validation. Taken together, the results from the two studies reported in this paper demonstrate, that artificial neural networks are a valuable and applicable alternative to classic algorithms of statistical judgment formation which can be used to considerably increase the precision of diagnostical decisions derived from test batteries. Unlike more classical methods of statistical judgment formation this new method also lends itself to a practically applicable selection of the most appropriate candidates based on their success probability in a relevant criterion measure such as flight simulators. Thus artificial neural networks constitute a decisive progress in pilot selection.

References

Ajzen, I. (1987). Attitudes, traits and actions: Dispositional prediction of behavior in personality and social psychology. In L. Berkowitz (Ed,), Advances in experimental social psychology (Vol. 20, pp. 1-63). New York: Academic Press.

Bortz, J. (1999). Statistik für Sozialwissenschaftler [Statistics for social scientists]. Berlin: Springer.

Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford: Oxford University Press.

Brown, M. T., & Tinsley, H. E. A. (1983). Discriminant analysis. Journal of Leisure Research, 15 (4), 290-310.

Brown, M. T., & Wicker, L. R. (2000). Discriminant analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), Handbook of applied multivariate statistics and mathematical modeling (pp.209-234). San Diego, CA: Academic Press.

Burke, E., Hobson, C. & Linksy, C. (1997). Large sample validations of three general predictors of pilot training success. International Journal of Aviation Psychology, 7, 225-234.

Damos, D. L. (1996). Pilot selection batteries: shortcomings and perspectives. International Journal of Aviation Psychology, 6, 199-209.

Dorffner, G. (1991). Konnektionismus [Connectionism]. Stuttgart: Teubner.

Goeters, K.-M. (1998). General standards of selection: Validity and utility analysis. In K.-M. Goeters (Ed.), Aviation Psychology: A science and a profession (pp.103-112). Aldershot: Ashgate.

Griffin, R. B. (1998). Predicting Naval Aviator Flight training Performance using multiple regression and artificial neural networks. International Journal of Aviation Psychology, 8, 121-135.

Hagemeister, C., Scholz, A. & Westhoff, K. (2002). Wie kann man Geübtheit in Konzentrationstests erkennbar machen? [How to unreveal routine in concentration tests?] Zeitschrift für Personalpsychologie, 2, 94-102.

Hunter, D. R. & Burke, E. F. (1994). Predicting aircraft pilot-training success: A meta-analysis of published research. International Journal of Aviation Psychology, 4, 297-313.

Kinnebrock, W. (1992). Neuronale Netze [Neural networks]. München: Oldenburg Verlag.

Lienert, G. A. & Raatz, U. (1998). Testaufbau und Testanalyse (6. Auflage) [Test construction and application (6th edition)]. Weinheim: Psychologie Verlags Union.

Martinussen, M. (1996). Psychological measures as predictors of pilot performance: A meta-analysis. International Journal of Aviation Psychology, 6, 1-20.

Masters, T. (1995). Advanced algorithms for neural networks. New York: John Wiley and Sons.

Mielke, A. (2001). Neuronale Netze [Neural networks]. [Online] URL: http://www.andreas-mielke.de/nn.html [01.10.2001].

Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1994). Machine learning, neural and statistical classification. New York, NY: Ellis Horwood.

Nabney, I. T. (2002). Netlab: Algorithms for pattern recognition. London: Springer.

Ree. M. J. & Carretta, T. R. (1996). Central role of g in military pilot selection. International Journal of Aviation Psychology, 6, 111-123.

Rojas, R. (2000). Neuronal Networks. A systematic introduction. Heidelberg: Springer.

Warner, B., & Misra, M. (1996). Understanding neural networks as statistical tools. The American Statistician, 50, 284-293.

Wittmann, W., & Süß, H.-M. (1997). Challenging G-mania in intelligence research: answers not given, due to questions not asked. Paper presented at the International Meeting of the International Society for the study of individual differences, 19-23 July, Aarhus, Denmark.

AIRPORT RESOURCE MANAGEMENT AND DECISION AIDS FOR AIRLINES

Amy Spencer
Cognitive Systems Engineering, Inc.
Columbus, Ohio

Philip J. Smith and **Charles Billings**
The Ohio State University
Columbus, Ohio

Airport surface management is fundamentally a task requiring decision making under uncertainty. For example, there is uncertainty about when an aircraft will be ready to push back, how long it will take a departing flight to taxi to the departure runway queue and how long it will take an arriving flight to taxi to its gate from the arrival runway. As a result, managing traffic on the airport surface, and coordinating this surface movement with airspace constraints, is a risk management task. Decision support tools which provide better access to airport surface data and predictions, as well as access to NAS-Status data such as airspace constraints, will reduce but not eliminate uncertainty. Therefore, to be effective, tools designed to support surface management decisions regarding events such as those listed above must reason about the inherent uncertainty in these events and assist airport users in their decisions regarding aircraft surface operations.

Introduction

In this paper we discuss the design and development of prototype tools developed to support NAS (National Airspace System) user decision making in regards to airport resource management and procedures that enable the effective use of these tools. These tools could not only increase operational efficiency on airport surfaces but could also result in a significant reduction in operational costs currently incurred by NAS users. Through the reduction of taxi times fuel burn can be significantly reduced; specific operator issues such as crew over-time, secondary de-icing, diversions and flight cancellations – which can be costly results of operational inefficiency on the airport surface – could also benefit from these types of tools.

As described in the Safe Flight 21 Pre–Investment Analysis Cost Benefit Analysis Phase II Report (May 1, 2001), Continental United States (CONUS) efficiency benefits for the introduction of better surface movement surveillance and planning tools were estimated to be $7.852 billion – two-thirds being attributed to "a reduction in taxi times as well as reduced arrival and departure delays". Another statistic that was quoted by a major cargo carrier indicated that a reduction in departure delay for their fleet by an average of 1 minute per year would save them $1 million per year.

The developed prototype was given the name ARMADA (Airport Resource Management and Decision Aid) and was developed as either stand alone software or to be integrated into the NAS user's existing software environment. Throughout the course of this work our objectives have included:

- Developing a concept of operation based on the use of programmable alerts to call the user's attention to important events that have been detected using airport surface data (integrated with other data sources as appropriate), and to provide active decision support for pushback planning decisions.
- Designing and implementing an interface design concept to demonstrate the nature and feasibility of this concept of operation and to identify important interface design features to enhance the usefulness and usability of these alerts.
- Developing an algorithm that provides predictions regarding the earliest that a runway queue is likely to run dry, and developing an interface to display this information to an airline Ramp Tower Administrator in order to support pushback decisions.
- Completing a formative evaluation of these prototype tools, eliciting input from prospective users regarding the potential usefulness, usability and value of the tools.

Our work in this area has been based on three fundamental premises:

- The availability and accuracy of technologies to provide real-time data on airport surface activity are reaching a point where they represent a viable source of information to improve airport operations.
- Uses of these data sources offer the potential to increase throughput on the airport surface as well as in the surrounding airspace. They also offer a means for NAS users to increase the efficiency and cost-effectiveness of their operations and they provide a means to enhance safety.

- In order to achieve these benefits in terms of throughput, efficiency and safety, it is not sufficient to provide only the NAS service provider (the FAA) with tools that access and make use of such airport surface data. NAS users must also have tools that make use of these surface data sources in order to plan and run their operations, and in order to coordinate effectively with FAA staff.

Users

Our focus has been on tools for organizations that support their flights with centralized operations centers and/or ramp control operations (including a significant number of General Aviation (GA) corporations and fractional ownership firms such as NetJet that make use of centralized operations centers to manage their flights), and that are therefore supported by specialists who, directly or indirectly, are helping to plan and coordinate the execution of airport surface activities. This means that our potential users are airline Dispatchers and Aircraft Routing Staff, ATC Coordinators, Ramp Tower Controllers, Ramp Tower Administrators, Gate Assignment Specialists, Gate Management Staff and Maintenance Staff. In addition to these direct users, the impact due to explicit communication or implicit coordination with other individuals will need to be considered, including FAA traffic managers and controllers at ATC Towers, Terminal Radar Approach Controls (TRACONs), Air Route Traffic Control Centers (ARTCCs) and the Air Traffic Control Systems Command Center (ATCSCC), as well as the crews of the affected flights. Our contention is that support of these user groups offers one of the major leverage points for increasing throughput and cost-effectiveness in the use of the NAS.

Approach

Our investigations have indicated that a mixed initiative interaction design is called for. In some cases, the user will recognize the need to look at a display to check for certain information. In others, however, the software needs to be monitoring for an important event and to alert the user about it. Our approach involved developing new display design concepts as well as algorithms that provide users with the information they need at the time they need it and in a form that they need it. We define these two areas as:

1. Programmable alerts and critiquing functions to support airport surface management.
2. Algorithmic support of pushback and sequencing decisions (using integrated airport surface and NAS-status information).

Due to space constraints, we will focus primarily on the design and development of programmable alerts and their associated displays.

In the remainder of this paper we will discuss the display design and functionality associated with the various ARMADA alerts and critiquing functions. There are numerous types of alerts that could be implemented and made operational very quickly (once a suitable design has been developed and appropriate surface data is made available at an airport). Some types of alerts would rely only on a combination of NAS user data (such as filed off time) and aircraft surface or terminal airspace positional data, while others would require the types of predictions generated by tools such as NASA's Surface Management System or SMS (Smith, et al., 2002). These alerts would not require any changes in current ATC practices and would be of use to all NAS users that make use of Ramp Control facilities at an airport and/or make use of a centralized operations center for dispatch functions.

Note: The distinction we make between alerts and critiquing functions is that critiques are a special type of alert made in response to some decision or action made by the user, rather than in response only to data input from the environment, while alerts in general can be triggered by external data and inferences made from these data. This requires special attention to the interaction design as, to provide a well-designed critiquing system, the interface between the user and the software must provide an unobtrusive source of data regarding the intentions or decisions of the user.

All of the various ARMADA alert displays have been designed to provide:

- Timely access to critical information including:
  - Actual and predicted OOOI times (OFF – departure time, ON – arrival time, OUT - gate push-back time and IN – gate parking time)
  - Inefficient operations or surface conflicts
- Access to context-sensitive detailed data displays upon demand. Display concepts include:
  - surface maps
  - airspace maps
  - timelines
  - tables (sortable)
- A communications function to support the efficient creation and sending of messages relevant to that alert
- Alert-specific user-customized parameters including:
  - Turning the alert on or off

- Determining which flight(s) or category of flight(s) to include in terms of a given alert
- Specifying alert timeframes (making the alert active only during specific times, such as during a departure push)
- Specifying how the alert will be presented (as a pop-up, or as some integrated display within an SMS display or an airline-specific display)
- Indicating the trigger(s) for the alert (location and/or flight status; spot involved; timeframe, etc.)
- Customizing the specific displays to include in the overall detailed display

Also, all ARMADA displays share certain general features including linked displays (if the user highlights an object in one display, information about that flight is highlighted in all displays where it appears) and a Find function (for finding and highlighting classes of objects in the display).

User Tasks

The various user tasks that the prototype tools were designed to support can be defined as either Departure Management, Arrival Management, Information Sharing/Coordination or Irregular Operations (Obradovich, et al., 1998; Smith, et al., 2002; Spencer, et al., 2003a; 2003b; 2002a; 2002b; 2002c; 2001). A comprehensive discussion of the tasks we have studied can be found in these papers.

Over the course of this work, we have conducted numerous studies including three site visits for data collection at the FedEx Ramp Tower and Global Operations Center in Memphis, and one site visit to Memphis ARTCC (ZME). These visits included the demonstration of interface designs and partial implementation of illustrative information displays and an algorithm that models uncertainty regarding taxi and departure times. We have also completed a formative evaluation providing data that is strongly supportive of the efficacy of our design concepts.

Interviews with Flight Operations, Ramp Tower and Dispatch staff at FedEx identified 12 tasks that these individuals thought would be of particular value.

- Delayed EDCT flights
- ESPed (Enroute Spacing Program) flights
- 18C/36C Runway departures
- Late Arrivals
- Spot Conflicts
- Gate Changes
- Long or Short Runway Queues
- Closed Routes (due to weather)
- Delays Associated with Deicing
- Runway Assignment Changes
- Diversions
- Pathfinder Selections

Based on this list, we selected four specific areas for concentration in our prototyping of alerts and associated displays:

- Late Arrivals
- Spot Conflicts
- Delayed EDCT flights
- 18C/36C Runway departures

These four areas were selected as they represent a range of different types of issues in terms of the underlying functionality and the required information displays. All four deal with performance by the Ramp Tower Administrator, but the general functionality applies to potential alerts for other airline staff as well. Due to space constraints we will limit our discussion below to details regarding alerts for late arrivals and spot conflicts.

Alert for Late Arrivals: Late arrivals can cause many different types of airline operational issues including issues regarding cargo or passenger connections, conflicts with departing flights (particularly in cases where the flight is arriving during a departure push and is therefore “traveling against the flow” of outgoing traffic), crew scheduling and gate assignment issues. This alert was designed to assist decision makers (who are often multi-tasking and working within a highly dynamic environment) in avoiding surface operations that may lead to these and other issues and in quickly finding appropriate solutions.

The Alert for Late Arrivals is designed specifically for the Ramp Tower Administrator (who oversees all ramp area operations, coordinates aircraft and surface vehicle movements and with FAA personnel, AOC staff and individual Ramp Tower Control positions as needed). As with all of the ARMADA alerts, this warning regarding a late arrival provides access to context-sensitive displays to aid in situation assessment and decision making. It is triggered whenever a late arrival reaches a certain state. As noted earlier, the user could set the various alert parameters to (for example) identify only certain flights, choose the flight state at which they want to be notified such as In Range, On Final or ON, and indicate how the alert should be presented - as a pop-up, or as some integrated display within an SMS display or an airline-specific display. If the user has chosen for alerts to be displayed within ARMADA then the first display is a pop-up alert window (see

Figure 1). This display contains critical information about this situation including the aircraft ID (ACID), flight status (In Range), predicted ON time, predicted IN time, assigned spot and parking gate. Note also that the interface allows the user to change the pre-set alert time (for example, the user can request to be alerted again when the flight is ON).

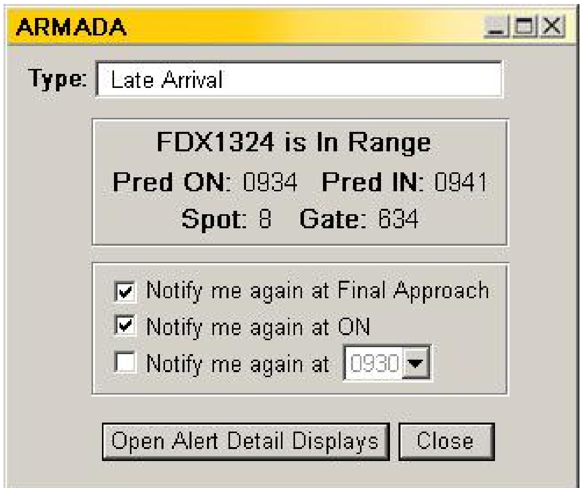


Figure 1. *Late Arrival Pop-up*

The pop-up interface also allows the user to temporarily minimize the alert and continue with other work for the time being (accessing it again later), close the alert, or choose to "Open Alert Detail Displays". Figure 2 is an example of the resultant displays if the user chose to view the alert detail displays.

Other functionalities related to the alert would enable quick access to other information, such as which flights should or could be held to prevent back ups (and extra waiting) in the ramp area, but would not explicitly include this information unless or until it was requested by the user.

In terms of the Alert for Late Arrival Detail Displays, note that:

- The inset map (upper left) is configured to show the airspace around the airport. In this display, we see that the late arrival has been enlarged on the map and is surrounded by a gold box. (In general, flights that ARMADA knows are directly involved in a situation triggering an alert are shown on the maps surrounded by gold boxes.)
- The more detailed surface map (upper right) shows all active flights as triangles color-coded by runway, and shows those flights with beacons (flights ready to push) as circular dots color coded by runway. Only those flights at the gates with beacons are shown as those are the departures that are still at their gates that could potentially interact with this arrival as they depart.
- The legend for the detailed map shows the numbers of active flights and flights with beacon by runway.
- The display subwindows are contained in a single larger alert detail display window so that they can be minimized or closed as a group.
- This display also provides functionality that allows the users to change their request for another later alert (at ON), and to close this alert permanently if desired.
- The Find function (upper left) was intended to allow the user to enter a specific aircraft ID or labels for categories such as "Heavy" aircraft or "ZNY" (New York ARTCC) departures and have the associated objects highlighted on the map.

Above we have indicated the proposed functionality and interaction design for the Late Arrival Alert. Overall, the response of the Ramp Tower Administrators and flight operations management to this approach, using programmable alerts to provide timely access to critical information, and providing access to context-sensitive detailed data displays upon demand, was extremely positive, both in terms of the potential usefulness and usability of our designs.

Spot Conflict Alert: Spot conflicts can cause considerable disruption to surface operations – it can take 45+ minutes to dispatch tugs to pull one of the involved aircraft out of the way. When these potential events are predicted there are several choices that an Administrator has: contact the FAA Tower and request that the arrival be held out of the ramp area or request that the arrival be brought in via a different spot, hold any involved departures at the gate or send any involved departures to a different spot. The time at which the potential conflict is detected determines, in part, what action the Ramp Tower Administrator may take. For example, if the arrival and possible spot conflict is detected at In Range, then the Ramp Tower Administrator would be more likely to direct the ramp controller(s) to move any departures to a different spot, or hold them at the gate. If the event involves an arrival and is detected at Final Approach, then the Ramp Tower Administrator would more likely contact the FAA Tower and ask them to hold the arrival out of the ramp area or request that they direct the arrival to a different spot. The reason for this is that if the potential spot conflict were not detected until Final Approach, then any involved departures would likely already be active.

Again, this alert is designed specifically for the Ramp Tower Administrator. Also, this alert is similar to the Alert for Late Arrivals in that it most likely deals with the unexpected event of an arriving flight attempting to enter the ramp area while departing aircraft are exiting the ramp area (or when a departing flight needs to return to its gate due to unexpected maintenance or other issues).

Note that, unlike the Alert for Late Arrivals which defines the relevant set of departures as those flights that are currently active or have a beacon (flights ready to push), the Spot Conflict Alert requires a more sophisticated set of predictions. In this sense, these two alerts illustrate our evolutionary approach. If the technology to predict spot conflicts is not yet mature enough, the Late Arrival Alert can be used to support the same user need (but requires additional assessments by the user) based on the information presented in the detailed display.

The Spot Conflict Alert requires acceptably accurate predictions of the taxi paths, spots and runways for departures and arrivals as well as predictions of the times associated with these different locations for a flight on the airport surface. This could be handled by using the deterministic modeling contained in SMS (based on fixed parameters for taxiway movement rates and departure rates), or by developing statistical models that use historical data to develop context-sensitive estimates of the uncertainty associated with different airport surface movements.

The Spot Conflict Alert displays have the same general features as all ARMADA alert detail displays.

Conclusion

Considerable evidence for the importance of this form of interaction, making use of software alerting functions, has been noted during our observational studies, structured interviews and focus groups. It is our conclusion based on our studies to date that many of the potential benefits from surface data will not be realized unless such alerting functions are developed to support the use of surface information by NAS users. It is not enough for the information displays in systems like SMS to be useful and usable when considered in isolation. The interaction design must be based on a realistic understanding of the operational demands of the user's environment (including all of the other tasks and information displays involved as part of his/her job).

Our conclusion is that this means that, in many cases, viewing of surface information needs to be supported on as "as needed" basis, with an alert triggering the user to check relevant surface information when some important situation arises.

This integrated, human-centered approach to the design of airport surface management decision support tools offers great potential as a strategy for enhancing the functioning of the NAS.

References

Federal Aviation Administration, Safe Flight 21 Pre–Investment Analysis Cost Benefit Analysis Phase II Report. 1 May 2001, Washington, DC.

Obradovich, J.H., Smith, P.J., Denning, R., Chapman, R., Billings, C., McCoy, E. and Woods, D. (1998). Cooperative Problem-Solving Challenges for the Movement of Aircraft on the Ground. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting,* 57-61.

Smith, P.J., Pear, C., Billings, C. and Spencer, A. (2002). *Support of Airline Operations Centers Using the Surface Management System*. Institute for Ergonomics Technical Report #2002-17. Ohio State University, Columbus OH.

Spencer, A., and Smith, P.J., (2003). Utility of the Surface Management System (SMS) within the FedEx Global Operations Center and Maintenance Operations Coordination Center. *Institute for Ergonomics Technical Report #2003-19*. The Ohio State University, Columbus OH

Spencer, A., Smith, P.J. and Billings, C. (2001). *Sample Scenarios Provided by an Airport Coordination Center at JFK*. Institute for Ergonomics Technical Report #2001-18. Ohio State University, Columbus OH.

Spencer, A., Smith, P.J. and Billings, C. (2002). *Ramp Control Issues in the Design of a Surface Management System*. Institute for Ergonomics Technical Report #2002-18. Ohio State University, Columbus OH.

Spencer, A., Smith, P.J. and Billings, C. (2002). *Surface Management System - Simulation2: Findings Regarding the TMC Position.* Institute for Ergonomics Technical Report #2002-19. Ohio State University, Columbus OH.

Spencer, A., Smith, P.J., Billings, C., Brinton, C. and Atkins, S. (2002). Support of traffic management coordinators in an airport air traffic control tower using a surface management system. In S. Chatty, J. Hansman and G. Boy (eds.), *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics*. Cambridge, MA: AAAI Press, 105-111.

Spencer, A., Smith, P.J., Billings, C., Brinton, C., Atkins, S. and Walton, D. (2003). Decision support and information exchange to improve airport surface management. *Proceedings of the 2003 Aviation Psychology Symposium*, Dayton, OH.

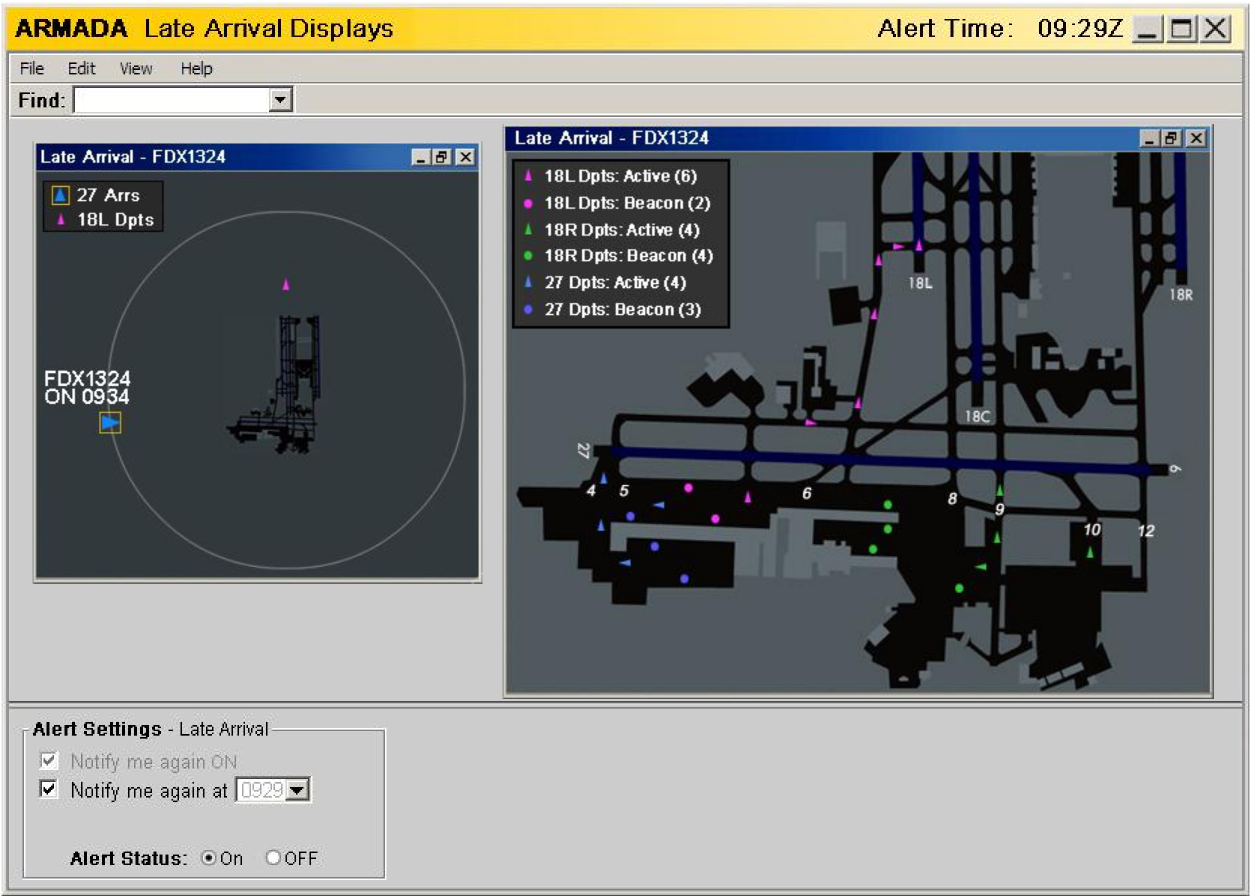


Figure 2. *Late Arrival Alert Detail Displays*

INFORMATION DISTRIBUTION TO IMPROVE TEAM PERFORMANCE IN MILITARY HELICOPTER OPERATIONS: AN EXPERIMENTAL STUDY

Brian K. Sperling and **Amy R. Pritchett**
School of Industrial and Systems Engineering
Georgia Institute of Technology, Atlanta, Georgia, USA

Changes in task requirements and system capabilities have led to the addition of crewmembers, information displays, and monitoring and coordination requirements in many domains. This experimental study tested the hypothesis that providing task relevant information to individual team members in a time critical environment, while limiting their access to non task-relevant information, would change team interactions by developing complementary team mental models and thus improve performance. The results of this experiment support this hypothesis, and give insight into how the distribution of information among team members effects the communications and coordination within a team. and team and individual performance.

Background and Introduction

The addition of a team member distributes cognition, changing the communication, coordination, and workload within the team (Hutchins and Klausen, 1996; Mosier et al., 2001). Although there has been significant research conducted on the display of information for a single operator and on group problem solving and performance (e.g., Orasanu and Salas, 1993), the current literature has identified the need to look more in depth at information and resource management within teams (e.g., Mosier and Skitka, 1996; Mosier, et al., 2001; Orasanu and Salas, 1993; Rouse, et al., 1992).

This research focuses on designing successful interactions between two team members in their naturalistic environment, in this case, a pilot and co-pilot in a military helicopter. This study hypothesized that providing specific task relevant information to individual team members in a time critical environment, while limiting their access to non-task relevant information, will change team coordination and assist in the establishment of complementary team mental models. Complementary team mental models are defined here as the condition in which:

- Each team member has the knowledge necessary to conduct his/her tasks.
- Each team member knows which information is known by the other team member should he/she need to seek it.
- Each team member knows which information is needed from them to other team members and when.

Historically, this type of team interaction knowledge is created by training, procedures, rules and regulations. In contrast, this study supports the concept that a "team centered" system design approach, focused on a complementary distribution of information among team members based on their tasks, will naturally promote improved team coordination by aiding team members in developing complementary team mental models. Furthermore, this method of distributing information among team members will provide individual crewmembers with a more accurate task relevant mental model of their environment.

The approach is somewhat counter-intuitive; traditionally the premise has been that increasing the amount of information that is shared between team members will naturally improve a team's shared mental model. This research supports the proposition that, in certain instances, a lesser amount of information overlap may improve a team's performance (e.g. Bolstad & Endsley, 1999). Similarity is a common gauge of effectiveness for team mental models. Yet this "similarity" often leads to inefficient team interactions. We propose the use of "complementariness", the mutual supplying of each other's lack, as a more reliable indicator of the efficiency of team mental models. The formulation of complementary team mental models can support team performance by helping to clarify roles and responsibilities, individual and team member information requirements, and improving the efficiency of explicit communications.

This study hypothesized that providing task relevant information to individual team members in a time critical environment, while limiting their access to non task-relevant information, would change team interactions by developing complementary team mental models and improve performance.

Method

Overview. This experiment was conducted at the US Army Aeromedical Research Laboratory (USAARL), located at Fort Rucker, Alabama using military helicopter pilots as participants. The main parameter was the complementariness of task specific information available to team members. During the experiment data were collected concerning team communications, crew workload, information requirements, decision-making and performance while the participants conducted a navigation task in a time critical situation. Each team member assumed a different role in the team, either the pilot-in-command (PIC) or the co-pilot/co-pilot (CPN); they maintained their assigned role throughout the entire experiment (i.e., there was no role switching).

Participants. Participants were 20 U.S. military rated aviators tested in pairs with the following characteristics:

- Their military rank ranged from Chief Warrant Officer II through Lieutenant Colonel.
- Participants' ages ranged from 24-57, with an average age of 39 years.
- Total flight hours ranged from: 210 to 11,180, with an average of 3290 hours.
- Each crew was required to have at least one crewmember rated in a dual engine aircraft.

Experiment Apparatus

NUH-60 Flight Simulator. The NUH-60 Black Hawk helicopter flight simulator used for this study was operated by a qualified simulator operator; see Figure 1. This provided an interactive environment in which team performance could be observed while certain parameters within the team were controlled.

Flight Instrument Cover-Ups. To force the division of information during the flight segments, the view of the instruments was blocked for the pilot and/or co-pilot. This was accomplished by physically obstructing the view of certain instruments in the cockpit with cardboard dividers attached with Velcro. Figure 1 shows information available to the co-pilot, but not the pilot, for example.

Foggles. Foggles are manufactured glasses used as a tool during instrument flight training. They limit the pilot's field of vision to the flight instruments.

Experimental Design

The study consisted of two experiments run sequentially; participants were unaware that there

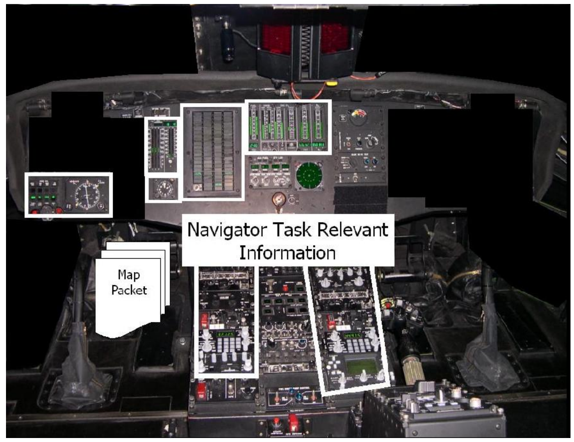


Figure 1. *NUH-60 Black Hawk Simulator*

were two experiments. The first experiment design consisted of two runs that examined performance under nominal conditions, during which the two levels of information distribution (complementary and normal) were varied. This experiment was balanced within subjects to account for order and training effects. The second experiment consisted of one experimental run similar to the previous two. However, the crew was required to deviate from normal procedures (i.e., react to an in-flight emergency). This was a balanced between subjects design between the two information distribution levels.

Scenarios

The flight profile incorporated various phases of flight during visual meteorological conditions (VMC) and instrument meteorological conditions (IMC). The profile has three sections to be flown in order, each lasting approximately 15 minutes. Flight phases of interest during VMC flight include take-off, VMC flight in cruise (above 200 ft AGL) and landing. Flight phases of interest during IMC flight include take-off, straight and level flight, climbs, descents, standard rate turns, and landing. All flight maneuvers were flown in accordance with Army standards.

Independent Factors

There were two independent factors in this experiment: complementariness of information and operational condition.

Complementariness of Information. Two levels of information distribution were presented to the participants: normal and complementary. A task analysis was used to determine the information each crewmember required access to in order to complete their individual and team tasks. Under the normal condition the pilot and co-pilot were both given identical information, i.e., they both had access to all information displays in the cockpit and they were both given a map with a route posted; they also were given a route card with headings, altitudes, airspeeds, and checkpoints, approach plates for local airfields, and a description of the landing area. Under the complementary condition individual crewmembers were only provided access to information relevant to their individual tasks and for their defined roles in team tasks. Specifically, only the co-pilot was given the map and relevant navigation information. Likewise, the co-pilot wore Foggles, preventing out-of-windscreen viewing. The pilot had access to all flight instruments but access to engine related performance instruments was restricted to the co-pilot. The co-pilot was not allowed to visually share the map, route card, etc. with the pilot.

Operational Condition. Two operational conditions were presented to the participants: nominal and off-nominal. During nominal conditions crews maintained visual flight rules throughout the simulation, and they experienced no system malfunctions during the mission. During off-nominal conditions crews experienced inadvertent instrument metrological conditions (IIMC) and a single engine alternator failure during the flight.

Dependant Factors

Data discussed in this paper were categorized into four main groups: performance, communication, workload, and information requirements.

Performance. During nominal flight conditions performance was measured by Flight Performance measures recorded by the simulator. Examples of these measures are: Root Mean Square Error (RMSE) of airspeed, altitude and heading, and rate of climb. Additional task performance measures were evaluated, including:

- *Completion of required radio calls*: Crews were given a list of radio calls required in each flight leg. This metric is represented by a percentage of those calls that were actually completed.
- *Calculation of estimated time enroute*: During each flight leg, co-pilots were required to calculate the estimated time enroute for two legs of each run.
- *Initiation of a fuel consumption check*: Crews were required to initiate a fuel consumption check during each run; this metric indicates whether or not this was completed.
- *Navigation and process errors*: Navigation errors concerned time, heading, distance, altitude, etc. Process errors include using the wrong frequencies, wrong procedures, etc.

During the off-nominal flight condition, in addition to the measures used in the nominal condition, the following task performance metrics were used.

- *Inadvertent instrument metrological conditions (IIMC) call time to Campbell Army Airfield (CAAF):* This was the time recorded from when the crew entered IMC until they notified CAAF.
- *Proper IIMC procedures*: A measurement of whether the crew performed the proper IIMC procedures in accordance with the aircrew-training manual.
- *Diagnosis time of emergency*: Time was recorded from the presentation of the emergency until the crew verbalized what the problem was or the corrective action needed.
- *Diagnosis of the proper emergency procedure*: This metric indicated whether the proper emergency procedure was executed.
- *Emergency call time to CAAF*: This was the time recorded from when the crew was presented with the emergency until they notified CAAF.

Communications. Verbal communications were categorized in three basic categories: transfers, requests, and acknowledgements (Entin and Entin, 2001) using the matrix in Figure 2. The data were normalized based on the length of each experimental run. Additionally, communication transfers were divided by communication requests to assess "anticipation ratio". Anticipation ratios have often proved more useful than individual rate measures for understanding team communications (Entin and Entin, 2001).

Type & Content		Navigator to Pilot	Pilot to Navigator	Total
Request	Task Relevant Information			
	Non Task Relevant Information			
	Action			
Transfers	Task Relevant Information			
	Non Task Relevant Information			
	Performing/ Will Perform Action			
Acknowledgements of Info Receipt	General (okay,roger)			
	Specific (roger...right on turn to 180 degrees)			

Figure 2. *Communication Matrix*

Workload. Workload was measured through the use of the NASA Task Load Index (TLX). Six sub-scales: mental demand, physical demand, temporal demand, performance, effort, and frustration. Workload was analyzed in two different manners. (1) Individual workload ratings were analyzed to determine whether individual crewmember's workload changed due to the cockpit configuration, and (2) a correlation analysis was performed using each crewmember's estimations of their teammate's significant sources of workload.

Information requirements. After each scenario, participants were asked to rank the importance of their information sources during each phase of flight, both for how important each type of information was to them and how important they believed it was for their crewmember. This was done for take-off, enroute navigation, and landing during nominal conditions, and upon entering IMC and dealing with the emergency procedure in off-nominal conditions.

Results

Performance

Nominal Flight Condition. The flight performance measures were analyzed using a General Linear Model (GLM) Analysis of Variance (ANOVA). Each task was evaluated based on predetermined parameter limitations. During nominal conditions no significant differences were found to exist due to changes in the distribution of information within the cockpit.

The task performance metrics were found not to fit the normality requirements for ANOVA. Therefore, each was examined using a Mann-Whitney test to identify the main effects of the independent variable. "Total errors" (p= .015) was significant; both process errors and navigation errors were marginally significant. Fewer errors were committed, in the nominal condition, when the information available was distributed in a complementary manner.

Off Nominal Flight Condition. The flight performance measures were analyzed using a GLM ANOVA. During entry into IIMC, the *RMSE for airspeed* was found to be significantly different as the cockpit configuration changed (p= .020). The median and mean RMSE decreased in the complementary condition; also the interquartile range of error in the complementary condition is less than the normal condition.

Across the complete flight profile for off nominal conditions the *percent of completed required radio calls* tended to increase when the crew was provided with complementary information; using the Mann-Whitney test, percent of radio calls was marginally significant with a p-value of .053. The Co-Pilot/Co-pilots (CPN) completed one hundred percent of the radio calls required during the complementary condition (see Figure 3). Both the median and mean percent of completed calls increased across conditions and performance was clearly more consistent during the complementary condition. Additionally, the *diagnosis time of aircraft emergency*, measured in seconds, could be analyzed using a GLM ANOVA and was found to have significant differences between the levels of information complementariness; the p-value was .007 with an observed power of .912. Figure 4 illustrates the direction of the difference and highlights the significant decrease in diagnosis time. Both the median and mean diagnosis time decreased across conditions from 130 to 49 seconds and 115.3 to 39.4 seconds respectively. Furthermore, the standard deviation decreased from approximately 34 to 20.

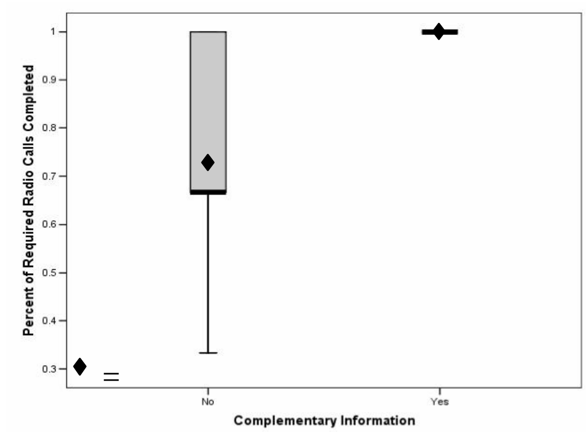


Figure 3. *Box Plot for Percent of Required Radio Calls Completed*

Communications

Nominal Flight Condition. A GLM ANOVA was used to evaluate all communication rates in the nominal condition (Figure 5). The following categories of team communication rates increased significantly when the crew was exposed to a complementary information distribution: Team Transfers of Action, Team Transfers of Task Relevant Information, Team Total Transfers, and Team Total Communications. Furthermore the rate of Transfer of Non-Task Relevant Information decreased in the complementary configuration, as did

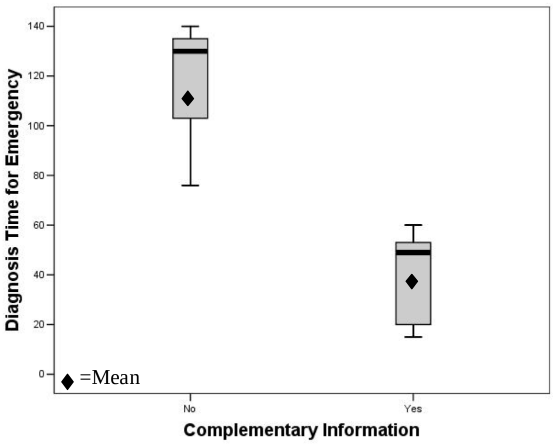


Figure 4. *Box Plot for Diagnosis Time of Aircraft Emergency*

the Team Anticipation Ratio (RTAR). Each detected change had a strong observed power; the lowest observed power was .842.

Off Nominal Flight Condition. A GLM ANOVA was also used to evaluate all communication rates in the off-nominal condition. The results in the off-nominal condition were very similar to the nominal condition. The following team communication rates increased significantly when the crew was exposed to a complementary information distribution: Team Transfers of Action, Team Transfers of Task Relevant Information Team Transfers of Task Relevant Information, Team Total Transfers, Team Acknowledgements Specific (RTAS), and Team Total Communications. In addition, The Team Anticipation Ratio decreased in the complementary configuration. Each detected change was accompanied by a strong observed power calculation; the lowest observed power was .657 (Figure 6).

Information Requirements

Rankings of information requirements from the pilot and co-pilot were matched by phase of flight and a correlation matrix was developed using the Spearman Rank Order Correlation Coefficient. Relevant correlations were analyzed using a GLM ANOVA for significant differences. The ANOVA performed for the nominal condition found no significant differences. On the contrary, differences in the mean correlation coefficients were significant in the off nominal condition due to changes in the complementariness of information in the cockpit (p-value = .0041); Figure 7 illustrates the increase in median and mean from the normal configuration to a complementary distribution of cockpit information.

Workload

Individual Workload Ratings were assessed using the NASA Task Load Index (TLX), and analyzed using a GLM ANOVA. Generally, there were no significant effects in team member's workload due to changes in the complementariness of information (see Table 1). The only significant change in mean ratings was detected in the co-pilots' mental workload in the nominal condition, which increased when operating in the complementary cockpit configuration. Additionally, there were four measures that were marginally significant; these measures also increase din the complementary configuration.

Workload Correlation. Crewmembers were also asked to estimate the sources of workload for their teammate using the modified NASA TLX scale.

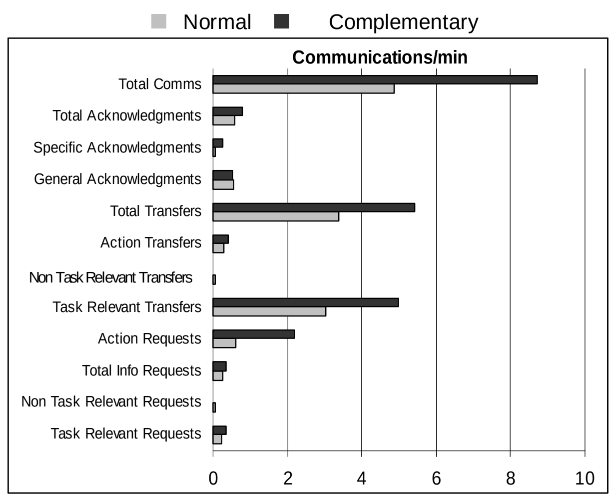


Figure 6. *Off Nominal Communications*

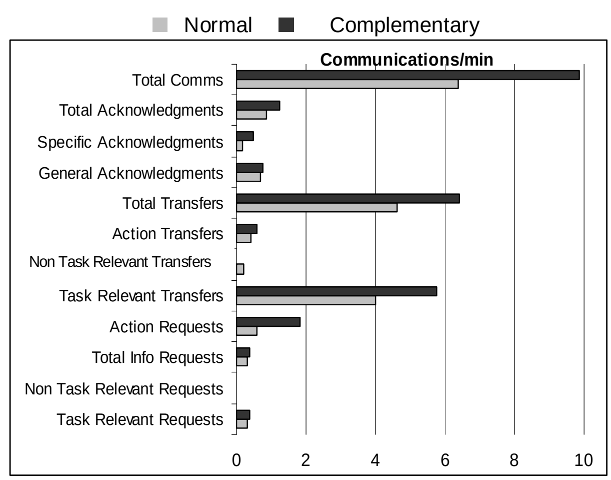


Figure 5. *Nominal Communications*

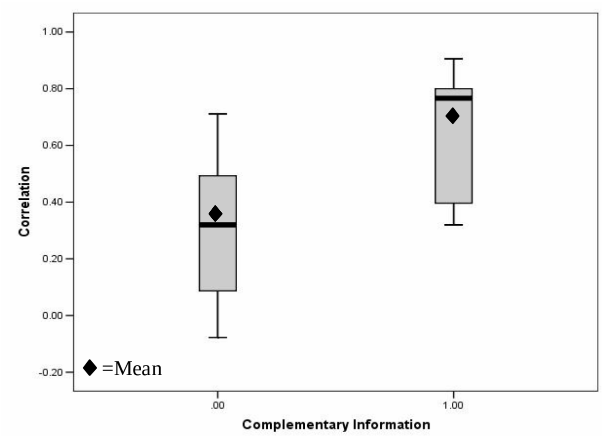


Figure 7. *Information Ranking Spearman Rank Correlation Coefficient*

Based on their scores, the sources of workload (mental, effort, temporal, etc.) were rank ordered (1-6). The pilot's ranking of the co-pilot's workload was matched with the co-pilot's ranking of the co-pilot's workload, and vice versa.

The ANOVA performed with the co-pilot as the found no significant effects. On the contrary, differences in the mean correlation coefficients were marginally significant (p = .069) when the pilot predicted the co-pilot's workload in the complementary configuration.

Conclusions

The results of this experiment provide empirical evidence that providing task relevant information to individual team members in a time critical environment, while limiting their access to non task-relevant information, improved individual and team performance by changing team interactions and helping to develop complementary team mental models. Furthermore, there is evidence of increased individual performance that indicates this method of distributing information among team members may provide individual

Table 1. *Individual Crewmember Workload Significance Levels*

	Nominal Helicopter		Off-Nominal Helicopter	
	Pilot	Navigator	Pilot	Navigator
Mental	0.700	**0.015**	0.765	0.407
Physical	0.694	0.657	0.925	*0.053*
Temporal	*0.086*	0.262	0.500	0.310
Effort	0.613	*0.068*	0.535	*0.054*
Performance	0.804	0.874	0.743	0.559
Frustration	0.165	0.432	0.266	0.380

crewmembers with a more accurate "task relevant" mental model of their own environment. The findings of this experiment give new insight into how the distribution of information among team members affects the development of shared expectations and information requirements, team and individual performance, and communications.

References

Bolstad, C. A., & Endsley, M. R. (1999) Shared mental models and shared displays: An empirical evaluation of team performance. In *Proceedings of the 43rd Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: Human Factors and Ergonomics Society

Entin, E.E., Entin, E.B., Measures for Evaluation of Team Process and Performance in Experiments and Exercises. *Proceedings of the 6th Int. Command and Control Research & Technology Symp*. June 19, 2001, Annapolis, MD.

Hutchins, Edwin & Klausen, Tove (1996): Distributed Cognition in an Airline Cockpit, in Engestrom, Y. & Middleton, D. (Editors): *Cognition and Communication at Work*, Cambridge University Press

Mosier, K., & Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other? In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc (pp.201-220).

Mosier, Skitka, Dunbar, McDonnell (2001). Aircrews and Automation Bias: The Advantages of Teamwork. *The International Journal of Aviation Psychology*, **11(1)**, 1-14.

Orasanu J., Salas E. (1993). Team Decision Making In Complex Environments. In Klein Associates, *Decision Making in Action: Models and Methods*

Rouse, W.B., Cannon-Bowers, J.A., & Salas, E. (1992). The Role Of Mental Models In Team Performance In Complex Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, **22 (6)**, 1296-1307.

THE ROLE OF INSTITUTIONAL REVIEW BOARDS IN AVIATION RESEARCH: IT'S THE LAW AND IT MAKES SENSE

Dr. Earl S. Stein
FAA, William J. Hughes Technical Center
Atlantic City International Airport, New Jersey

Research in medicine and social sciences often involves the participation of human participants, who under the rules in place today volunteer their time and understand both the benefits and risks associated with the research. This was not always the case. Rules, regulations, and laws currently require oversight by organizations referred to as Institutional Review Boards (IRBs). These boards exist to protect the participants, ensure their ethical treatment, and encourage good research. IRBs enhance the quality of research planning, and the IRB process should be part of every researcher's timeline for completion of his/her projects.

Research involves a systematic search for a reality that transcends our concepts as individuals. While philosophers will debate that there are many realities, in science we attempt to narrow the options. In social science we usually state our conclusions in probabilistic terms, admitting that there is some chance we could be wrong.

We base our conclusions on data gathered from the systematic study of some phenomenon such as behavior. We have and still study the behavior of animals and then make comparative assumptions about how their conduct may mirror our own actions. In some cases this is necessary, because it would be considered unreasonable or unethical to conduct certain studies with human beings. However, such ethics or rules of scientific morality have not always been followed and under some socio political conditions they have been ignored entirely in the misguided belief that science transcends all.

We collectively tend to forget about the "good old days" when researchers could pretty much do whatever they wanted in the name of science. There was no oversight and no IRBs. Those were the days when humans could be put at risk without knowing what the risks were, or in some cases that they were even participating in a research project. Most researchers followed their professional ethical codes and remained within the scope of law at the time. Some did not. Many walked the fine line in between. This led to notable examples which made the media in the 50's and 60's because of disastrous results.

There are many citations concerning research gone too far. The sources, themselves, can sound at times like reactionary paranoia from anti-research or anti-government organizations. For example, Smith (1998) noted "since World War II, the United States Government, mainly the Central Intelligence Agency, has secretly and at times inhumanely sought a way to control human behavior"(p. 1). Dr Frank Olsen, a Department of Defense employee, was a notable example of the CIA's LSD research program. He was given LSD without informed or any other consent; it led to depression and his suicide (Elliston, 2004). The US Army also experimented with LSD and a psychoactive gas, quinuclindnyl benzilate (BZ), from 1955 to 1975 at Edgewood Arsenal Maryland, on soldier "volunteers", who were told they would experience transitory discomfort and could terminate the experiment any time they wished but only with the consent of the physician in charge (Edgewood Guinea Pigs, 2004). This was not exactly informed consent as we know it today.

Other organizations also conducted experiments that today we would likely find unacceptable. Universities participated under grant or contract relationships with the government. In 1977, testifying before a Senate committee Admiral Stansfield Turner, then director of the CIA, admitted that his agency has participated in research involving drugs and other "mind" altering methods (Turner, 1977). While this work took place before he became director, he agreed to notify all living participants but debated about notifying participating universities in that public knowledge of the work could damage their reputations.

This is not to say that this work went on with no ethical code or rules in place. They did exist but were somehow overlooked or set aside, no doubt in part under the premise of national security. The National commission for the protection of Human Subjects was established by the National Research Act in 1974. The Tuskegee Syphilis Study was one of the factors that helped create this law.

In the Tuskegee Syphilis study. poor African American men with the disease were left untreated so researchers could follow the progress of the disease. They were not informed volunteers. The following quote is from the Centers for Disease Control Website:

"The Tuskegee Syphilis Study, carried out in Macon County, Alabama, from 1932 to 1972, is an example of medical research gone wrong. The United States Public Health Service, in trying to learn more about syphilis and justify treatment programs for blacks, withheld adequate treatment from a group of poor black men who had the disease, causing needless pain and suffering for the men and their loved ones" (CDC, 2005, p. 1).

In part to help comply with the National Research Act, the Department of Health Education and Welfare commissioned a group of researchers and ethicists to meet at the Belmont Conference Center of the Smithsonian Institution. Their mission was to define the ethical principles and guidelines necessary for future human based research (NIH, 1979). The Belmont report summarizes the key ethical principles that the commission identified.

This work grew out of the Nuremberg code, which evolved from the trials of the same name, and was originally a method of judging physicians and other scientists who participated in research during World War Two. The conferees noted that ethics is all about boundaries and what constitutes reasonable behavior as compared to that which is deemed unethical.

The authors of the Belmont report made a clear distinction between research and practice in both medical and behavioral research. Practice involves interventions designed to improve the condition or well being of a patient or client. Research is about testing hypotheses, drawing conclusions, and advancing the body of knowledge. If research and practice occur in the same setting, or if there is any doubt as to whether research is an element in the overall program, human review for the protection of participants is required.

There are three general principles around which research ethics should be based: respect for persons, beneficence and justice.

Respect for persons is an acknowledgement that each individual is autonomous and has a right to consent or not. Part of this is to determine whether the individual has the ability to understand and if in diminished capacity extra protection is required. *Beneficence* is a principle that infers as researchers we should do no harm and both maximize the benefits and minimize the risks associated with the research. This may require a balancing of the potential rewards of doing the research against the potential risks to participants. The last principle is *justice*. Do members of the population have an equal chance of being selected for participation or does the burden of participation fall on a subgroup based on who they are or how much they have? According to the American Psychological Association (APA) (2002) in their outline of the ethical principles for psychologists, "justice" implies that psychologists ensure to their best efforts that everyone can benefit from the processes, procedures, and services they offer. As well, they must avoid the impact of their own biases and their own limitations in competence and experience so that unjust practices (i.e. the Tuskegee Syphilis Study) do not occur ever again.

The three general ethical principles are implemented through application in research. *Informed consent* is the application of respect for persons. APA calls this the respect for people's rights and dignity or Principle E. According to the Belmont report, informed consent has three parts: information, comprehension and voluntariness.

Information is provided which is accurate and sufficient so that a "reasonable volunteer" can clearly understand the risks and benefits. Incomplete disclosure is only allowed if complete information would bias or materially change the study, all risks are still disclosed, and there is a plan for debriefing participants after the data is collected.

Comprehension is the second key element. Information is provided in a manner and pace that facilitates understanding and if necessary, the researcher is obligated to test for comprehension either verbally or in writing. The third element is *voluntariness*. Participation must be truly voluntary and not coerced in any way. The research cited from Edgewood Arsenal where participants could only leave with permission did not begin to meet that criterion. We would also not want to see the type of influence that researchers can have as found by Stanley Milgrim (1974) in his work on obedience to authority. Deception was used and no aftercare plan for participants was conceived or implemented. The main lesson that came out of Milgrim's work was that ordinary people would do extraordinary things given the right social pressures in an environment labeled as research.

The Belmont conferees noted that the second application of the principles involves the assessment of risks and benefits. This is based on beneficence. Is the study worth doing given the potential outcomes weighted against the actual risks for participants? By risks they mean more than a probability but the nature and extent of harm that could befall a participant. These include both the psychological and the physical. A review committee can also consider the long term benefits of the research that may go beyond those for the individual participant and the costs of not doing the research and the loss of those benefits.

The University of Michigan Medical Institutional Review Board website (2004) commented as follows on the Belmont report:

> "The Belmont Report, as monumental as it may be, did not make specific recommendations for administrative action by the Secretary of Health, Education and Welfare; rather, it recommended that the report be adopted in its entirety, as a statement of the Department's Policy. What dignity, what statesmanship! The Belmont Report laid three basic ethical principles: "Respect for persons. Beneficence. Justice." Respect for persons; beneficence; justice. How simple, how fundamental, how awesome; not just for research involving human subjects, but for everything we do every day."

While the Belmont report was basically an outline with recommendations, the rules it recommends are codified in Federal Law (DHHS, 1983). Under 45 CFR 46 the guidelines for use of human subjects (participants) are specified and the role of Institutional Review Boards is defined. The Department of Transportation is covered specifically under 49 CFR 11 and this is a word for word copy of the DHHS regulation. The regulation clarifies what constitutes research, whether or not human beings are research subjects and also notes that even if 45 CFR 46 does not apply, other Federal, state and local laws may come into play.

Recently the Office of Human Research Protections, which is part of DHHS, published a series of decision charts designed to assist researchers and Institutional Review Boards in making decisions concerning Research proposals. Figure 1 is presented as an example (DHHS, 2004 September).

The Federal regulations and laws apply to all research funded by the, or accomplished within the Federal government. Other state and Federal laws may apply as well. Further, most professions involved in human research have ethical codes which in some ways are as stringent as Federal Law. Those of us in Psychology adhere to the APA Ethical Code or one similar to it. In the Federal Aviation Administration we have FAA Order 9500/25 which essentially mirrors 45 CFR 46 up to subparagraph 124 then goes on to offer additional protections for other specified subgroups of potential populations, such as prisoners with whom FAA researchers generally do not work (DOT, 2004). These regulations require the existence and operation of Institutional Review Boards or IRBs.

The IRB is where the researcher using human participants (note the not so subtle change from "subject" which is the term most regulations use) meets the Institutional requirements as specified in law and regulations. Many researchers including this author have at one time or another viewed the IRB by whatever title (i.e. peer review committee in Universities) as basically an impediment, a roadblock, and other terms, some even stronger, to imply that IRBs hold them up and ask them to do unreasonable things. A number of authors writing about IRBs have commented that in addition to evaluating participant safety and confidentiality IRBs should evaluate what would be lost or the cost of not doing the research that they may disapprove (Rosnow, Rotheram-Borus, Ceci, Blanck, Koocher, 1993; Rosenthal, 1994). The Belmont report had implied this as well.

IRBs are made up of people who are in many ways very much like the folks who must staff their research plans with the boards. The laws and regulations specify the general membership of an IRB. Each board must have at least five members of varied backgrounds. It can not consist of only members of one profession. The board can not be all men or women. It must include at least one member whose primary interests are in science and one member whose interests are outside of science. Members may not review research proposals in which they may have a conflict of interest.

The FAA's rules for membership are even more specific than those of the Federal Law: (1) One member who is a physician, with clinical experience or specialization in aerospace medicine. (2) One member with expertise in the behavioral and social sciences. (3) One member who is not an employee of FAA, with expertise in ethics. (4) One member with expertise in safety or industrial hygiene (in addition to review of research protocols, this member also shall, at the direction of the IRB Chair, conduct on-site inspections to assess overall safety of the proposed research projects). 5) One member representing the FAA Chief Counsel.

Currently the FAA has two IRBs. The primary IRB, which covers the entire FAA, is based in Oklahoma City. There is also a local IRB which operates at the

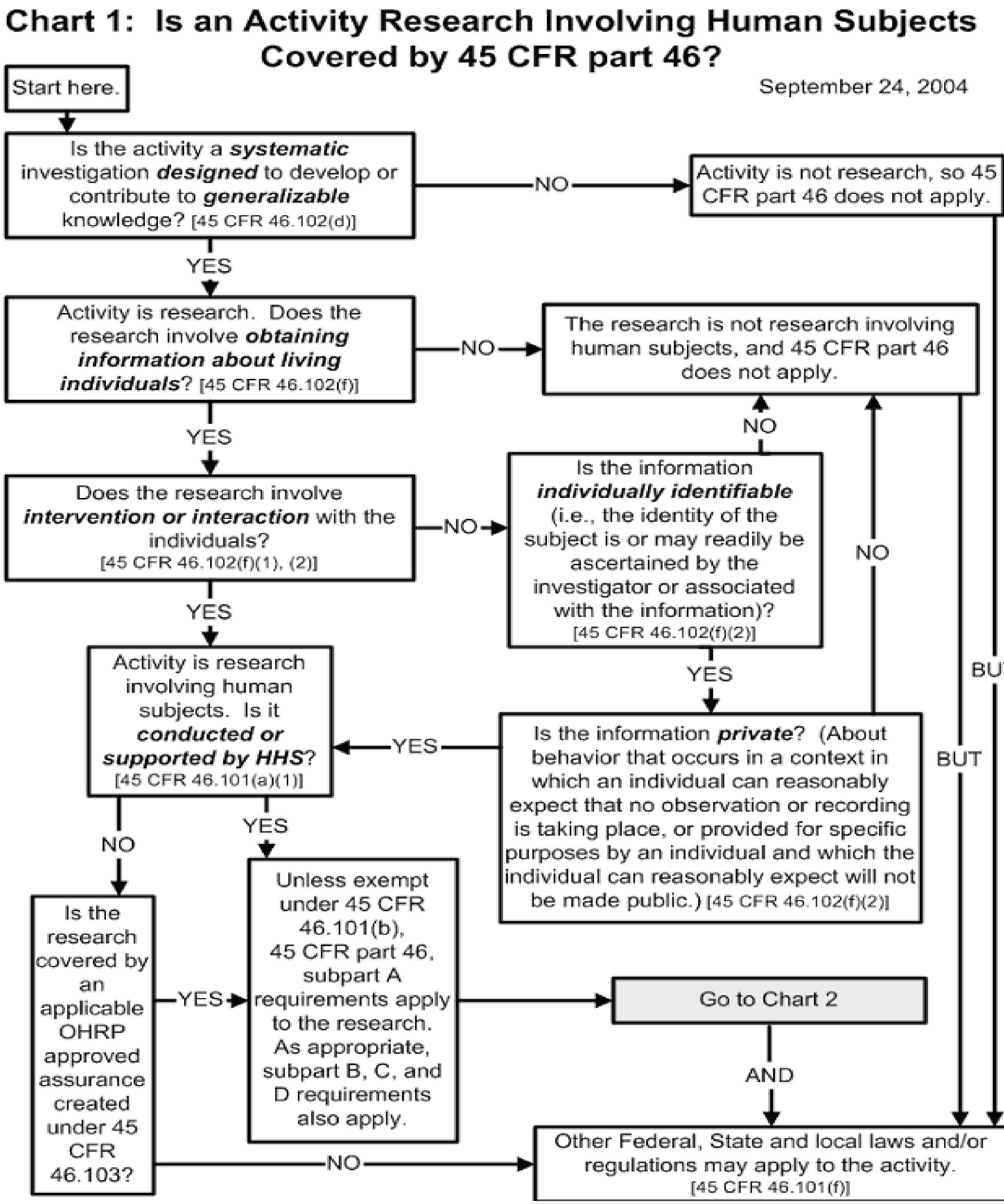


Figure 1. *Decision Support Chart*

FAA Technical Center. The local IRB handles only those research proposals that fit under minimal or no risk standards. Fortunately, this covers most of the research done at or for the Technical Center. The local IRB has a membership which meets all of the legal and regulatory requirements specified above. The physician is a local private practice internist who is a certified flight surgeon. The ethicist who is not directly affiliated with the FAA is a Chaplain with the New Jersey Air National Guard Wing based at Atlantic City International Airport.

One criticism of IRBs in general is that they are inconsistent. Rosnow et al. (1993) reported that one research plan was approved by an IRB at one university and disapproved by another university in the same community. Sure, this can happen. At least within the FAA IRBs, we are all following the same regulation with the same intent of not stopping research but rather promoting better, ethically based, and well planned research.

The purpose of the IRBs is not and was never to impede good research. IRBs are there to ensure the safety of participants and verify that a volunteer is a volunteer who really knows what he or she is getting into and knows what the risks are. The IRB is also there to ask the question, "Are the risks worth the benefits of the research?" IRB members are encouraged to ask what would be lost if the research was not conducted.

The existence of IRBs encourages (some might say forces) researchers to plan carefully and to use planning tools such as check lists to avoid missing some key points in the planning process. For example, do they intend to sample from a special population such as children or prisoners that require additional protections and scrutiny? We do not see this much or at all in the FAA. However, the plan has to have an informed consent statement and agreement that is clear and well written. If it does not, we do send it back, even if informed consent is described in the body of the plan.

This is not done to annoy the researchers. They did have a copy of the guidelines and checklist, which forms the cover sheet on our local board's application package. Further, the IRB process encourages the researcher to know the population from which he or she is sampling, so that they are reasonably certain when someone agrees to participate, informed consent is truly informed and not an attempt to please the researcher.

IRBs are not enforcement organizations. They exist to provide a means for researchers to comply with the law and regulations. It is up to management within Federal organizations and the FAA in particular to enforce the adherence to the requirements. If managers and researchers do not comply, they risk sanctions if something should go wrong in a study, and they have not followed the rules in preparation for the research. The key is to plan so that the probability that things go wrong is low and a reasonable person would not have foreseen the problem as likely to occur.

There are a number of advantages for researchers to not only accept the IRB process as a fact of their research lives but to embrace it. It allows them to comply with the law and regulations. It increases the probability that all bases are covered so that the level of risk or lack thereof, they believe exists, is in fact the level of risk present during the study. This protects the institution and the individual researcher. It ensures that the research is being done in an ethical way and participants know what they are getting into when they grant informed consent. These are definitely good results. Yes, the IRB adds time to the planning process for a study, but you can include that in your overall plan. It should not be a surprise to anyone.

References

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist* 57, 1060-1073.

Center for Disease Control. (CDC). (2005). *The Tuskegee Syphilis Study: A hard lesson learned*. Retrieved January 8, 2005 from http://www.cdc.gov/nchstp/od/tuskegee/time.htm

Edgewood guinea pigs: Covert US Army medical experiments on human test subjects. (2004). Retrieved October 1, 2004 from http://www.parascope.com/gallery/galleryitems/edgewood/edgewood.htm

Elliston, J. (Ed.). (2004). *MKULTRA: CIA mind control*. Retrieved October 1, 2004 from http://www.parascope.com/ds/mkultra0.htm

Milgrim, S. (1974). *Obedience to authority*. New York: Harper.

National Institute of Health (NIH). (1979, April). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research, the National Commission for the protection of human subjects of biomedical and behavioral research.* Retrieved December 9, 2004 from http://ohsr.od.nih.gov/guidelines/belmont.html

Rosenthal, R. (1994). Science and ethics in conducting, analyzing and reporting psychological research. *Psychological Science*, 5, 127-133.

Rosnow, R.L., Rotheram-Borus, M. J. Ceci, S. J., Blanck, P.D. & Koocher, G.P. (1993, July). The institutional review board as a mirror of scientific standards. *American Psychologist*, 48(7), 821-826.

Smith, A. (1998). *Mind control, LSD, the CIA and the American people: What the government does not what you to know.* Retrieved October 1, 2004 from http://www.mindcontrolforums.com/lsd-mc-cia.htm

Turner, S. (1977). *Project MKULTRA, the CIA's program of research in behavior modification-CIA director Standfield Turner's testimony*. Retrieved October 1, 2004 from http://www.druglibrary.org/schaffer/history/e1950/mkultra/hearing05.htm

U.S. Department of Health and Human Services (DHHS). (1983). Protection on human subjects. *Code of Federal Regulations* 45 CFR 46.115. Washington, DC: US Govt Printing Office.

U.S. Department of Health and Human Services (DHHS). (2004, September). *Human subject regulations decision charts*. Retrieved November 15, 2004 from http://www.hhs.gov/ohrp/humansubjects/guidance/decisioncharts.htm

U.S. Department of Transportation (DOT). (2004). Protection of human subjects. *Code of Federal Regulations* 49 CFR 11. Washington, DC: US Govt Printing Office.

University of Michigan Medical Institutional Review Board. (2004*). An overview of ethics, laws and regulations regarding participation in research.* Retrieved December 9, 2004 from http://www.med.umich.edu/irbmed/volunteers/ethics.html

DISCUSSING MONOTONY IN ATC: EFFECTS OF REPETITIVE TRAFFIC PATTERNS ON PERFORMANCE AND SUBJECTIVE INDICATORS

Sonja Straussberger
Karl-Franzens-University – EEC
Brétigny sur Orge, France

Wolfgang Kallus
Department of Psychology, Karl-Franzens-University
Graz, Austria

Dirk Schaefer
Eurocontrol Experimental Centre (EEC)
Brétigny sur Orge, France

This study addresses the concept of monotony in ATC and describes uneventful and repetitive work conditions evoking such a state. Psychophysiological effects of repetitiveness in two simulated ATC-scenarios of low or high dynamic density (DD) were investigated with 24 air traffic controllers (M=29.5 years, 18 male, six female). Interactions approached significance ($p<0.1$) in conflict resolution time for an out-of-routine conflict situation. Conflict resolution lasted longer in repetitive traffic and resolution time increased from the first to the second run. Those findings are supported by a composite score of subjective attentiveness, fatigue, sleepiness, and concentration with lower values found in repetitive conditions. Although generally decreasing, the switch from low to high DD was rated favorably on the hedonic tone, while tense arousal was reacting more pronounced in non repetitive and low-high condition. In combination with the development of earlier reported cardiovascular (heart rate, heart rate variability) and subjective indicators the results underline the significance of a multidimensional assessment of monotony in ATC.

Introduction

This paper discusses the concept of monotony in air traffic control (ATC) and reports results of an experimental study. Hopkin (1995, p. 341ff.) presents a number of arguments in favor of a more thorough investigation of monotony and boredom in ATC. Amongst these arguments are controllers' complaints and presumed performance impairments as well as long-term consequences expressed in job satisfaction. Indeed, research has advanced little on that subject within the domain of ATC.

Furthermore, the ambiguous and sometimes unreflected use of terms like monotony, boredom, low vigilance, underload and even fatigue has been addressed by various authors (e.g., Davies, Shackleton & Parasuraman, 1983; Thackray, 1981). This has not yet resulted in a clear distinction and goes beyond the scope of this paper. Nevertheless, it is noted that from a historical point of view apparently different priorities had been set by European and North American researchers resulting in a separate development and domination of theories addressing monotony, boredom or vigilance. The International Standard for mental workload (ISO 10075) recommends handling low vigilance and monotony as independent task attributes. Since the main purpose of the standards is to provide practitioners with guidelines of work design, the offered distinction does not provide a satisfactory theoretical basis.

A comprehensive theory of the concept of monotony included therein was offered by Bartenwerfer (1960, 1985). He stated that monotony is a specific consequence in working situations when continuous engagement in a task is required. Those tasks are of a restrictive nature and may be characterized by low stimulus intensity or variation, high repetition, low difficulty level and longer time on task. As a consequence, physiological deactivation and self-reported feelings like boredom, tiredness or sleepiness are registered. Increased reaction times and reduced ability to readapt after changes characterize performance impairments. This concept seems to be a suitable background for further research in ATC as it stresses the multidimensional effects of predefined task demands. As early as the beginning of the last century, job monotony became a subject of scientific interest, predominantly with the goal to optimize work performance amongst assembly line workers. Another aspect of monotony became relevant when Mackworth (1948) started to study performance of military control personnel to explain failures in signal detection. This contributed to a long tradition of research under the concept of vigilance.

Unfortunately, the long tradition in research on monotony has ignored that totally different task characteristics might lead to complaints about monotony. This was finally considered by Johansson (1989) who distinguished *uneventful* and *repetitive* monotony. Her distinction focuses mainly on control-room operators as an example for uneventful monotony compared to assembly line work representing repetitive monotony. Apparently, in her paper she also adopts the term monotony to describe stimulus conditions. This maintains the unclear classification of monotony as cause and consequence. In contrast, Bartenwerfer (1985) emphasized differentiating objective monotonous working conditions as a cause for an individual state of monotony. Following this statement, the authors prefer to use the term *repetitiveness* (respectively *uneventful* work conditions) to address task characteristics and monotony to indicate the individual response.

Johansson's distinction facilitates the systematization within the concept as it does not a priori exclude the vigilance concept. However, it has not yet been considered that uneventful and repetitive monotony might occur within one job. Such an example is represented by Air Traffic Control. Rather obviously, uneventful monotony can emerge in situations of low traffic that require few actions. Such a situation varies for regions and centers, but often occurs during night shifts. This aspect has been addressed within vigilance research (e.g., Schroeder et al., 1994) with the argument that such monitoring situations mainly demand sustained attention. Nonetheless, results of vigilance research are not directly transferable to the state of monotony. Focusing on this component neglects the complex nature of ATC. As task analysis revealed (Redding, 1994; Kallus et al., 1999), major task processes not only comprise monitoring, but checking, diagnosing and decision making complete the action cycle. The balance of these components might be different, but is still present even in conditions of very low traffic. In addition, the continuous update of a controller's mental representation of the situation requires active behavior, e.g. frequent scanning of the situation or communication with colleagues. Repetitive monotony can result in medium or high traffic conditions if task characteristics do not display a certain variation or if difficulty remains below a challenging threshold. Under these conditions, the nature of traffic has the potential to cause monotony in many centers, as repetitiveness can be found in various traffic conditions. Examples include runway allocation affecting approach and departure routes, certain sector forms, routine traffic, or parallel airways with few crossing points.

The framework of Johansson contributes to a better description of various aspects of monotony, but needs to be adapted to the working conditions within ATC (e.g., complexity, predictability, work environment, payment, and amount of control) that are used to differentiate the work on assembly-lines and in control rooms and are not directly transferable to ATC. Even in situations of low traffic, a certain complexity is available thus the action cycle includes a variety of steps for task execution. Conversely, errors resulting from both types of monotony might have different reasons. While in uneventful monotony they might occur because of suboptimal activation and consequently slow re-adaptation, in repetitive monotony errors might emerge out of routine that causes omissions in the update of the action cycle. A difference in the psychophysiological monotony pattern need not be shown in either situation.

Few studies have been conducted to better describe monotony in ATC. Thackray et al. (1975) were interested in physiological and subjective changes accompanying monotony and boredom. They found that the group reporting high monotony and boredom showed greater increases in response times, HRV, and strain while attentiveness decreased. In a field study Hoffmann and Lenert (1993) administered a questionnaire with the scope to assess strain reactions in controllers. Increased subjective monotony and fatigue were found towards the end of the shift. Traffic complexity counteracted this effect.

To summarize, there is a need to systematically investigate conditions which cause monotony in ATC considering individual and situational factors. The present study was designed to investigate the role of repetitiveness in simulated ATC. It was hypothesized that physiological and subjective state as well as performance will change due to repetitiveness in traffic characteristics. Furthermore, an influence of traffic complexity was assumed. A simplified version of the dynamic density (DD) concept was introduced (e.g., Laudeman et al., 1998) because it allows an appropriate description of the developing traffic situation over time.

First results from cardiovascular indicators (heart rate, heart rate variability) and subjective ratings have already been reported by Straussberger, Kallus & Schaefer (2004). The repetitive traffic condition resulted in physiological deactivation (decreasing HR, increasing heart rate variability). Mean values in feeling of monotony revealed higher ratings for repetitive scenarios but were interacting with the sequence of Dynamic Density changing from high to low versus low to high.

The present report focuses on performance components and includes subjective ratings.

Method

Participants

Twenty-four fully qualified air traffic controllers (18 male, six female) of Maastricht Upper Area Control Centre (MUAC) individually volunteered in this study. The session was performed during their planned working schedule. Age ranged from 22 to 47 years (M= 29.5, SD=6.0), on the average they had been fully licensed for six years (SD=5.5) and originated from ten European nations. Participants were randomly assigned to the experimental groups and did not differ in age or professional experience.

Experimental Design

Independent Variables. The experiment involved a 2 x 2 x 2 x 2 - mixed design. The between-subject-variables comprised repetitiveness (repetitive vs. non repetitive traffic pattern) and sequence of dynamic density (high-low vs. low-high). Each participant completed two scenarios (run 1 vs. run 2), the second within-factor concerned the *intervals within runs* and was included if repeated measurements were analyzed.

Dependent variables. To determine performance during scenarios, conflict resolution times and the number of Short Term Conflict Alert (STCA) events of an out-of-routine conflict situation at the end of the scenario were extracted from simulator log files. They contained information about aircraft position, STCAs and controller actions. The measurement of resolution time started from the appearance of the aircraft in conflict until the first action (change in FL or heading) was undertaken.

Subjective ratings of attentiveness, fatigue (inv.), concentration and sleepiness (inv.) were collected on a 7-point-scale (1=low; 7=high) each 15 minutes until the end of the scenario. After level-corrections those items were summarized in an indicator to reflect how efficient participants felt during performing. Bipolar mood dimensions were assessed with the UWIST Mood Adjective List (UMACL; Matthews et al. 1990) and included tense arousal (anxiety vs. calmness), energetic arousal (vigor vs. tiredness) and hedonic tone (contentment vs. depression).

Other Variables. Confounding variables of interest were boredom proneness, action control strategy, initial well-being, initial stress-recovery-state, and personality traits. Control variables also comprised biographic data. Further effects of an introduced countermeasure and additional physiological measures (EEG, EOG, and EDA) will be reported separately. A detailed description of cardiovascular and subjective measures (further ratings included strain, boredom and irritation, feeling of monotony, NASA-TLX) can be found in Straussberger et al. (2004).

Procedure

A separate simulation room was allocated for the study on the premises of MUAC. Participants were allocated either to the morning session at 8:00 or to the afternoon session at 14:00, counterbalanced for experimental conditions. Before participating in the session, they were provided with information and signed a consent form. An average session lasted 5.25 hours. The experimental session started with 90 minutes of briefing, preparation for physiological recordings and set-up familiarization. After 15 minutes of rest break, two traffic scenarios of 45 minutes each were run. The introduction of the countermeasure required the completion of a short third run. Physiological recordings were collected with a Vitaport III recorder (Temec Inc.) throughout the session, including several baselines at the beginning and end of the scenarios. Subjective ratings were collected during the traffic scenarios. UMACL and other questionnaires were administered subsequently. Participants were video-taped during scenarios and a debriefing concluded the session.

Task

Participants worked on a simulated controller working position (CWP) including a 28'' LCD monitor with keyboard and mouse for inputs; STCA was available and Reduced Vertical Separation Minimum (RVSM) for Europe applicable. The simulation environment involved a semi-generic upper airspace (FL 250 – FL600) created for this experiment that was implemented as a standalone sector with two automatic feed sectors (no pseudo-pilots). The sector involved arriving and departing traffic from a major airport.

Four traffic scenarios with medium traffic load (57 aircraft per hour) were created according to the experimental manipulations. Regularly occurring potential conflicts would result in a very close near-miss in three-minute-intervals if the controller did not take appropriate action.

In repetitive scenarios, participants were presented with potential conflicts occurring at the same crossing point. This situation consisted of an aircraft in departure meeting an incoming northbound aircraft after two minutes in the sector. The non repetitive scenarios contained potential conflicts at varying crossing points throughout the sector. In order to obtain a task-performance-indicator, an out-of-routine conflict situation was introduced in the last interval of the scenario.

For the manipulation of DD, major factors such as number of aircraft, number of level changes, routes, and crossing points remained constant in three-minute-intervals throughout the scenario. The manipulation between high and low DD was implemented with additionally required level changes.

Controllers were instructed to control traffic as usual; a deviation concerned keeping aircraft on the planned route without redirecting.

Results

Statistical analysis employed a repeated measure ANOVA with repetitiveness and sequence of DD as between factors, and run and intervals within run as within factors. An alpha level of .05 was used for statistical tests. Differences in course and effects of DD were determined from trend analysis. It is noted that any significant interactions between sequence of DD and run express effects of counterbalancing.

Table 1. *Mean conflict resolution time (SD) for repetitive and non repetitive traffic and l-h vs. h-l sequence of DD with n=23*

Repetiti-veness	Repetitive		Non repetitive	
DD	**l-h**	**h-l**	**l-h**	**h-l**
Run 1	279.83 (141.78)	287.17 (104.30)	294.83 (41.35)	153.80 (62.44)
Run 2	301.83 (86.39)	305.00 (66.40)	280.33 (110.46)	315.60 (40.13)

Table 1 displays mean and standard deviation of conflict resolution time. One subject was excluded from analysis as conflict resolution time could not be determined in one run. ANOVA did not reveal significant differences for the main factors.

There is a tendencially significant increase in conflict resolution time from the first to the second run (F1=3.69, p=.070). Interactions between run and sequence of DD (F1=3.12, p=.093) and between run, sequence of DD and repetitiveness (F1=3.43, p=.080) are approaching significance and depicted in Figure 1.

Table 2. *Frequency of STCA events (STCA/ No STCA) for out-of-routine conflict situation (n=24, 2 Scenarios)*

	Low DD	High DD	Total
Repetitive traffic	3/9	3/9	**6 / 18**
Non repetitive traffic	0/12	2/10	**2/22**
Total	**3/21**	**5/19**	**8/40**

The frequency of STCA (Table 2) that occurred in the out-of-routine conflict situation represented a very rare event. For this reason, the factor run was excluded from analysis and DD (low vs. high) treated as between subjects variable. The Exact Fisher Test was run separately for each factor to examine the distributions of STCA events compared to no STCA events and resulted in no significant difference for either variable.

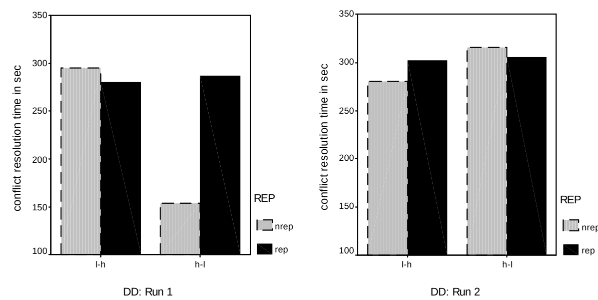


Figure 1. *Conflict resolution time in the first and second run depending on repetitiveness and dynamic density*

The indicator for subjectively reported feelings of efficiency revealed a main effect of repetitiveness (F1=9.80, p=.005), with lower ratings during repetitive traffic. An overall decrease occurred from the first to the second run (F1=23.16, p=.000) and within one scenario (F2=37.31, p=.000). Significant interactions were found between interval and repetitiveness (F2=6.76, p=.003, linear F1=13.03, p=.002). A significant interaction between run, interval and repetitiveness is depicted in Figure 2 (F2=5.49, p=.008; linear F1=1.97, p=.012). Also the interaction between repetitiveness, sequence of DD, run, and interval resulted in significance (F2=4.01, p=.026; linear F1=5.51, p=.019).

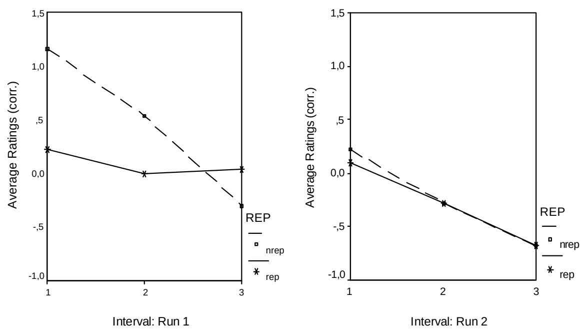


Figure 2. *Mean composite score of attentiveness, concentration, sleepiness, and fatigue (level corrected) during first and second run for repetitive and non-repetitive traffic conditions (REP=repetitiveness, nrep= non repetitive, rep= repetitive) for n=24*

Subjective mood was assessed on three dimensions (Table 3). On the subscale hedonic tone a significant main effect of sequence of DD was found. Participants rated their hedonic tone significantly higher (F1=6.68, p=.017) when they executed the scenarios in the order from low to high DD, but a general decrease from the first to the second run was found (F1=6.93, p=.016).

Table 3. *Average ratings (SD) for mood dimensions (HT=hedonic tone, TA= tense arousal, EA= energetic arousal) depending on repetitiveness (repetitive vs. non repetitive traffic) and sequence of DD (h-l vs. l-h) for n=24*

	Repetitive-ness	**Repetitive**		**Non repetitive**	
	DD	**l-h**	**h-l**	**l-h**	**h-l**
HT	**Run 1**	2.94 (.32)	2.81 (.21)	2.94 (.32)	2.77 (.33)
	Run 2	2.79 (.19)	2.60 (.18)	2.89 (.20)	2.52 (.18)
TA	**Run 1**	2.92 (.19)	2.90 (.18)	2.56 (.38)	2.92 (.17)
	Run 2	2.98 (.22)	2.73 (.18)	2.69 (.30)	3.00 (.22)
EA	**Run 1**	1.56 (.17)	1.67 (.19)	1.73 (.48)	1.65 (.09)
	Run 2	1.63 (.24)	1.44 (.10)	1.75 (.14)	1.63 (.31)

Tense arousal revealed significant interactions between repetitiveness and sequence of DD (F1=6.39, p=.020), between repetitiveness and run (F1=5.83, p=.025) and sequence of DD and run (F1=4.38, p=.049). In the conditions of non repetitive traffic respectively low-high sequence of DD tense arousal increased from the first to the second run. Average values were generally higher for repetitive and high-low condition, a slight decrease occurred from the first to the second run. No significant differences emerged on the energetic arousal subscale.

Discussion

The importance of monotony has been underestimated in ATC and this subject has not yet been well researched. The present study attempts to contribute towards filling this gap. In the introduction we outlined that it is important to consider both uneventful and repetitive work conditions as potential causes of a state of monotony. We focused on the latter and investigated it in simulated ATC. The reported results on performance and the subjective dimension complement those of physiological and other indicators described in Straussberger et al. (2004). In our previous work the physiological deactivation pattern was found in cardiovascular indicators and accompanied by increased ratings of feeling of monotony in repetitive traffic conditions. The current analyses support the results as subjectively perceived impairments were found in an indicator summarizing attentiveness, concentration, fatigue, and sleepiness.

But the multidimensional assessment of a state of monotony as proposed by Bartenwerfer also considers impairments on the performance level. For this reason, the conflict resolution time and frequency of STCA events in an out-of-routine conflict situation where studied. We found that conflict resolution time increased from the first to the second run and was longer in repetitive conditions. Although statistically not significant, the distributions of STCAs complete this picture. Low mean values found in the group that performed the first run in non repetitive high DD conditions are not caused by individual outliers. Furthermore, the values in the repetitive traffic condition demonstrate a wider range.

The decrease in hedonic tone expresses that the traffic density and its sequence affect the controllers' well-being. Even though descriptive values indicated decreases in repetitive conditions on energetic arousal, its insignificant result might have been influenced by manipulations in DD to result.

To a certain extend, the results can be compared to those of Thackray et al. (1975), as they found a similar cardiovascular pattern for the group with high ratings in feeling of monotony and boredom. Also, they rated their attentiveness lower and showed performance impairments. While their interpretation focused on reduced attention, we prefer to explain the results with a general impairment in the individual state. An advantage of the present study is that the sample consists of air traffic controllers and the simulation environment offered a better representation of reality.

The data also indicate that monotony develops rather soon, whereas after a longer time-on-task general fatigue overlaps with consequences of repetitiveness.

Nonetheless it is surprising that a state of monotony can result in ATC as a probable consequence of repetitive traffic conditions, especially since up to date research focused predominantly on situations of stress and vigilance.

Conclusion and Outlook

Future analysis will address the influence of confounding variables and result in recommendations concerning work organization and selection of controllers. Nevertheless it should be kept in mind, that further studies in the field will be necessary to better understand the origin of these phenomena.

It is critical that developments in ATC do not ignore research on monotony. This is especially true due on the one hand to an ongoing trend towards automation, and on the other hand in consideration of controllers handling increasingly complex traffic in the future. In both cases, the role of monotony needs to be clearly addressed and understood, as it is implicitly included in many attempts to cope with predicted traffic increases.

References

Bartenwerfer, H. (1960). Untersuchungen zum Monotonieproblem [Studies in the field of monotony]. *Zentralblatt fuer Arbeitswissenschaft*, 14, 29-33.

Bartenwerfer, H. (1985). *Monotonie in unserer Arbeitswelt – muss das sein?[Monotony at work – is this necessary?]* Muenchen: Bayrisches Staatsministerium für Arbeit und Sozialordnung.

Davies, V. J., Shackleton, D. R. & Parasuraman, R. (1983). Monotony and Boredom. In: G. R. J. Hockey (Ed.), *Stress and Fatigue in Human Performance* (pp. 1-32). New York, Wiley.

ISO 10075 „Ergonomic principles related to mental workload", Part 1: ISO 10075-1:1991 General terms and definitions, Part 2: ISO 10075-2:1996 Design principles, Part 3: ISO 10075-3: 2004 Principles and requirements concerning methods for measuring and assessing mental workload .

Johansson, G. (1989). Job demands and stress reactions in repetitive and uneventful monotony at work. *International Journal of Health Services*, 19(2), 367-377.

Kallus, K. W., Frazens, K., Barbarino, M., Vandamme, D. & Dittman, A. (1999). Integrated task analysis of en-route control: a process oriented approach. In: Jensen R. et al. (Eds.). *10th International Symposium on Aviation Psychology* (pp. 541–547). The Ohio State University.

Laudeman, I. V., Shelden, S. G. Branstrom, R. & Brasil, C. L.(1998). *Dynamic Density: An Air Traffic Management Metric*, NASA-TM-1998-112226.

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, 1, 6-21.

Matthews, G., Dylan, J. & Chamberlain, A. G. (1990). Refining the measurement of mood: The UWIST Mood Adjective Checklist. *British Journal of Psychology*, 81, 17- 42.

Redding, R. (1992). Analysis of operational errors in Air traffic control. *Proceedings of the Human Factors Society Annual Meeting*, Vol. 2, Santa Monica, CA: Human Factors Society, 1321-1325.

Redding, R. E. & Seamster, T. L. (1994). Cognitive Task Analysis in air traffic controller and aviation crew training. In: N. Johnston, N. McDonald & R. Feller (Eds.). *Aviation psychology in practice* (pp. 190 - 222). Hants: Avebury Technical.

Schroeder, D. J., Touchstone, R. M., Stern, J. A., Stoliarov, N. & Thackray, R. I. (1994). *Maintaining Vigilance on a Simulated ATC Monitoring Task Across Repeated Sessions* (DOT/FAA/AM 94/6). Washington, DC: Department of Transportation /Federal Aviation Administration, Office of Aviation Medicine.

Straussberger, S., Schaefer, D. & Kallus, W. (2004). A psychophysiological investigation of the concept of monotony in ATC: Effects of traffic repetitiveness and traffic density. In: *Proceedings of 1st International Conference on Research in Air Transportation (ICRAT)*, University of Zilina, Slovakia, 22-24 November 2004, 199-208.

Thackray, R. I. (1981). *Boredom and monotony as a consequence of automation: A consideration of the evidence relating boredom and monotony to stress* (FAA-AM-81-1). Washington, DC: Department of Transportation/Federal Aviation Administration, Office of Aviation Medicine.

Thackray, R. I., Bailey, J. P., & Touchstone, R. M. (1975). *Physiological, subjective, and performance correlates of reported boredom and monotony while performing a simulated radar control task* (FAA-AM-75-8). Washington, DC: Department of Transportation/Federal Aviation Administration, Office of Aviation Medicine.

Acknowledgements

The PhD study of Sonja Straussberger is funded by Eurocontrol Experimental Centre in Brétigny sur Orge, France.

HUMAN PERFORMANCE MODELLING FOR ACCIDENT RISK ASSESSMENT OF ACTIVE RUNWAY CROSSING OPERATION

Sybert H. Stroeve, **Henk A.P. Blom**
National Aerospace Laboratory NLR
Amsterdam, The Netherlands

A human performance modelling approach is presented for risk assessment of operations with multiple, dynamically interacting agents. The approach is illustrated for a risk model of runway incursion on an active departure runway. This model-based approach can provide detailed, systematically derived results on risk contributions of human operators and technical systems in complex multi-agent environments.

Introduction

Since capacity and efficiency are the drivers of the development of advanced air traffic operations, by now there is a broad consensus that appropriate accident risk assessment models are needed to assess safety in relation to capacity with the aim to optimise advanced air traffic operations (Wickens et al., 1998). Air traffic operations account for highly distributed and dynamic interactions between human operators, procedures and technical systems. As such, the safety of air traffic operations depends not only on the functioning of the individual elements in such multi-agent scenarios, but also on their complex interactions, especially in non-nominal situations. Because of this distributed control nature of air traffic, established techniques fall short in performing accident risk assessment. Blom et al. (2001) addressed this problem by development of a Monte Carlo simulation-based methodology that takes an integral approach towards human performance modelling and accident risk assessment for air traffic (Traffic Organization and Perturbation AnalyZer: TOPAZ).

The human performance modelling approach followed in TOPAZ is based on a contextual perspective in which human actions are the product of human internal states, strategies and the environment (Amalberti and Wioland, 1997; Hollnagel, 1993; Wickens and Holland, 1999; Cacciabue, 1998). The model for task performance of a human operator considers multiple tasks, human error and contextual control modes (Blom et al., 2003). Specifically, for a human operator

- a decomposition of the tasks of the human operator is identified,
- the most essential cognitive control modes are identified,
- the characteristics of the operator tasks are identified for the most important cognitive control modes,
- clusters of tasks are identified,
- hierarchy and concurrency for the task clusters are identified.

In such performance modelling, parameter values are based on operational observation, real-time simulation and expert interviews. Corker et al. (2005) showed that an additional way of identifying parameter values is to make use of the more detailed human performance model of Air-MIDAS.

In air traffic, situation awareness problems are important contributing factors to many accidents. The concept of situation awareness addresses perception of elements in the environment, their interpretation and the projection of the future status (Endsley, 1995). In an air traffic environment with multiple human operators, these aspects and associated errors of situation awareness depend on human-human and human-machine interactions. A model for situation awareness evolution in a multi-agent air traffic environment was developed (Stroeve et al., 2003; Blom and Stroeve, 2004). Here, an agent is an entity, such as a human operator or a technical system, which may have situation awareness of its environment. The environment of an agent includes the complete group of agents. The situation awareness of each agent consists of time-dependent information of other agents, including identity, continuous state variables, mode variables and intent variables. Achieving, acquiring and maintaining situation awareness depends on processes as observation, communication and reasoning.

It is the goal of the current paper to elucidate the approach for multi-agent human performance modelling and illustrate it for simulation-based accident risk assessment of an active runway crossing operation. In the sequel of this paper, the risk assessment steps and the runway operation are described first, followed by methods and results of the simulation model with emphasis on human performance aspects.

Accident Risk Assessment Steps

Following the TOPAZ methodology, assessment of the risk of an operation is performed in a number of steps:

1. Determine the scope: In collaboration with operational experts, determine the scope of the operation. Determine safety criteria and methods of the risk assessment.
2. Description of the operation: Describe in sufficient detail the operation, including context, human roles and responsibilities, procedures and technical systems.
3. Hazard identification: Identify non-nominal events or situations possibly having adverse effects on the operation. Particularly of interest are brainstorm results on situations and events for which pilots and controllers have complementary opinions.
4. Construction of conflict scenarios: Hazards are related to conflict types and ordered with respect to time and cause and effect. The resulting hazard structures are called conflict scenarios. Risk is divided into sub-risks related to the various conflict types. This enables efficient and orderly evaluation of risk.
5. Argumentation-based evaluation: Evaluate the risk based on the conflict scenarios, interviews with operational experts (pilots, controllers) and incident databases. This provides a first indication of the severity and frequency of conflict scenarios.
6. Development of a simulation model: Develop a mathematical accident risk model for conflict scenarios that are difficult to assess by argumentation-based evaluation. This stochastic dynamic model represents the performance and interaction of technical systems and human performance for a particular air traffic situation.
7. Simulation-based evaluation: Evaluate potentially safety-critical and uncertain risks by Monte Carlo simulations based on the developed simulation model and hierarchical simulation speed-up techniques.
8. Evaluation of model assumptions: Assess the effect on the modelled risk of assumptions made in the modelling process. This step accounts for the recognition that a model differs by definition from reality. It includes an analysis of bias and uncertainty in assumptions as well as a risk sensitivity analysis, and results in an evaluation of bias and uncertainty bounds of the risk of the operation.
9. Risk criteria: Compare the evaluated risk with risk criteria to assist decision-makers in their evaluation of the acceptability of the operation.

Here, operational experts are actively involved during hazard identification, argumentation-based evaluation and evaluation of model assumptions.

Active Runway Crossing Operation

The active runway crossing operation enables traffic to cross an active departure runway (named Runway A) in order to taxi between the aprons and a second runway (named Runway B). Each crossing has remotely controlled stopbars on both sides of the runway. The operation includes a large number of interacting agents (see also Figure 1):

- aircraft (taking off or taxiing),
- aircraft's flight management systems (FMS),
- pilots flying (PF's),
- pilots not flying (PNF's),
- Runway A controller,
- Runway B controller,
- ground controller,
- departure controller,
- start-up controller,
- ATC system, which is broadly defined to include
  - airport manoeuvre control systems,
  - surveillance systems,
  - airport configuration,
  - environmental conditions,
  - communication systems.

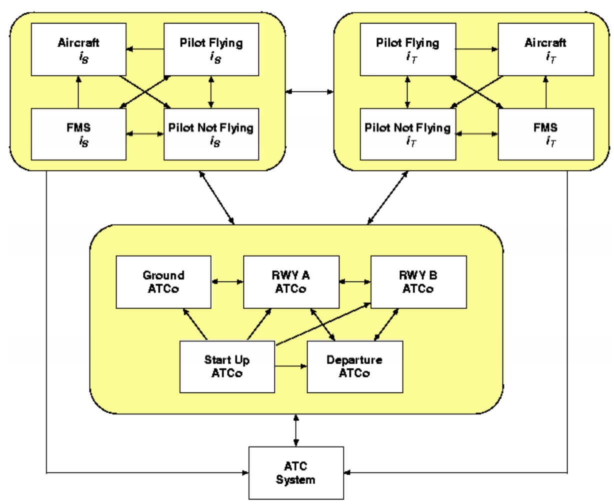


Figure 1. *Relations between agents identified for the active runway crossing operation.*

In the operation, communication between controllers and aircraft crews is via standard R/T. Monitoring by the controllers is via direct visual observation and is supported by radar track plots. The runway crossing operation over Runway A is under the responsibility of the Runway A controller. The Runway A controller is supported by a runway incursion alert system and a stopbar violation alert system. The Runway A controller manages the remotely controlled stopbars and the runway lighting. Monitoring by the aircraft crews is by visual observation and may be supported by the VHF R/T party-line effect.

Simulation Model

An initial argumentation-based evaluation of the risk of the active runway crossing operation showed that of all identified conflict scenarios, there are three conflict scenarios that may pose unacceptable safety effects. In this paper, we focus on the details of an accident risk model for one of these conflict scenarios. In this conflict scenario there is one aircraft that takes off and has been allowed to do so and there is one aircraft that crosses the runway while it should not. Taxiing along a straight line over a standard runway crossing is considered. Hence, in the illustrative example of this paper, emphasis is placed on the models of the aircraft, pilot flying, Runway A controller and ATC system agents. A high-level overview of these models is specified next.

Aircraft A taking-off aircraft initiates take-off from a position at the beginning of the runway. A crossing aircraft initiates crossing at a position close to the remotely controlled stopbar with a normal taxiing speed or from a hold state.

Pilot Flying of Taking Off Aircraft Initially, the pilot flying (PF) of a taking off aircraft has the SA that take-off is allowed and initiates a take-off. During the take-off the PF monitors the traffic situation on the runway visually and via the VHF communication channel. The PF starts a collision avoidance braking action if a crossing aircraft is observed within a critical distance from the runway centre-line or in reaction to an ATCo clearance, and it is decided that braking will stop the aircraft in front of the crossing aircraft.

Pilot Flying of Crossing Aircraft Initially, the PF has the intent SA that the next airport way-point is either a regular taxiway or a runway crossing. In the former case the PF proceeds taxiing and in the latter case the PF may have the SA that crossing is allowed. The characteristics of the visual monitoring process of the PF depend on the intent SA. In case of awareness of a conflict, either due to own visual observation or due to an ATCo call, the PF stops the aircraft, unless it is already within a critical distance from the runway centre-line.

Runway Controller The Runway A controller visually monitors the traffic and has support from a stopbar violation alert and a runway incursion alert. If the ATCo is aware that a crossing aircraft has passed the stopbar, a hold clearance is specified to both the crossing and the taking off aircraft.

ATC System The ATC system includes communication systems, tracking systems, a stopbar violation alert, a runway incursion alert and remotely controlled stopbars.

Hazard Representation The model of the active runway crossing procedure accounts for intent-dependent and cognitive mode-dependent error-prone perception processes of pilots flying and the Runway A controller. Table 1 shows how a number of situation awareness related hazards of the operation considered were accounted for in the accident risk model.

Table 1. *Examples of the representation of hazards in the accident risk model of the active runway crossing procedure.*

Hazard	Model representation
Runway incursion alert is active, but runway controller has wrong ‘picture’ of the situation, and therefore reacts too late, not or wrongly.	In response to an alert there is a chance that the runway controller does not observe the conflict and therefore does not react.
Pilots get confused because of complexity of the taxiways in the new operation.	The PF of a taxiing aircraft may be aware that the aircraft is taxiing on a regular taxiway while it actually is on a runway crossing.
Pilot reacts not, wrongly, too late or cannot react to conflict solving clearance of runway controller.	There is a chance that the PF does not or only after a long time becomes aware of a clearance.

Performance Model of Pilot Flying

The various human performance submodels are integrated into a simulation model. As an illustrative example, a model is presented of the pilot flying of an aircraft that taxies towards the runway crossing. A high-level overview of the model elements of the pilot flying agent is shown in Figure 2. The human operator model includes the following groups of model elements.

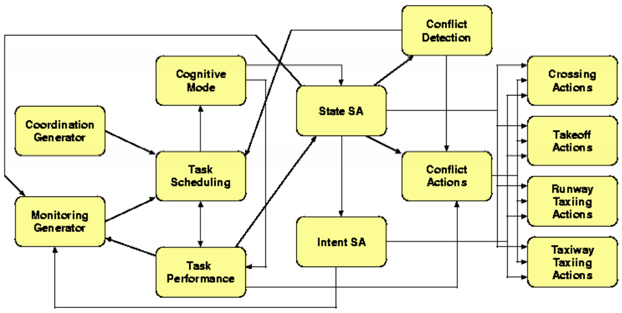


Figure 2. *High-level overview of the model elements of the pilot flying agent.*

Task Triggering Task triggering processes specify times at which it is desired to complete a task. They may depend on other processes, such as task performance and situation awareness. For example, the model blocks Monitoring Generator and Coordination Generator in Figure 2 represent task triggering processes of a pilot flying and specify times at which monitoring of the traffic situation and coordination with the pilot not flying is desired, respectively. These model blocks receive several inputs. For instance, the dependence of Monitoring Generator from Intent SA enables an intent-dependent visual updating frequency.

Task Scheduling Task scheduling processes determine which tasks should currently be processed by the human operator. Task scheduling processes may depend on other processes, e.g., task triggering, task performance and situation awareness processes. For example, in Figure 2 the Task Scheduling block represents a scheduling process with a fixed hierarchy and concurrency structure.

Task Performance Task performance processes describe the development of the progress of a task. They may, e.g., depend on task scheduling and cognitive mode processes. For example, in Figure 2 Task Performance depends on Cognitive Mode, resulting in a faster task performance in the opportunistic control mode with respect to the tactical control mode of the pilot flying.

Cognitive Control Mode Cognitive control mode processes describe the cognitive control mode of the human operator. They may, e.g., depend on the number and types of scheduled tasks. See, for instance, the Cognitive Mode block in Figure 2.

Situation Awareness Situation awareness model elements represent the state SA and intent SA, as outlined before. In Figure 2, the model blocks State SA, Intent SA and Conflict Detection represent SA components, where the latter block represents the detection process and the SA of a conflict. In Figure 2, State SA depends Cognitive Mode, representing that (errors in) the state SA updating process can depend on the cognitive mode.

Task Specific Actions Task specific actions represent particular elements of tasks of a human operator. For instance, for a pilot flying these may include (see Figure 2) Crossing Actions, Takeoff Actions, Runway Taxiing Actions, Taxiway Taxiing Actions and Conflict Actions.

Hierarchical Monte Carlo Simulations

An accident risk assessment includes a risk decomposition, which supports efficient evaluation of the collision risk and promotes insight in the risk contributions. The evaluation of the collision risk is based on the probabilities and the conditional collision risks of combinations of event sequences, as have been identified in the decomposition process. The decomposition process considers whether alert systems, remotely controlled stopbar and communication systems are functioning well or not. The decomposition process considered in the example includes

- the aircraft type of each aircraft to be either a medium-weight A320 or a heavy-weight B747;
- the intent SA of the PF of a crossing aircraft concerning the next way-point (Taxiway / Crossing) and concerning allowance of runway crossing (Allowed / Not Allowed);
- whether alert systems are functioning well or not;
- whether the remotely controlled stopbar is functioning well or not;
- whether communication systems are functioning well or not.

Based on the simulation model and the accident risk decomposition, Monte Carlo simulation software is developed to evaluate the conditional collision risk for the events resulting form the decomposition process.

Accident Risk Results of the Model

This section presents results of the simulation-based risk evaluation for a generic runway in good visibility conditions. Figure 3 shows the accident risk as function of the distance of the runway crossing with respect to the runway threshold. The probability of a collision decreases for larger crossing distances. Figure 3 also shows the decomposition of the total risk for the cases that the pilot flying of the taxiing aircraft either intends to proceed on a normal taxiway (without being aware to be heading to a runway crossing) or intends to cross the runway (without being aware that crossing is currently not allowed). The largest contribution to the risk is from the situation that the pilot intends to proceed on a normal taxiway. The relative size of this contribution depends on the crossing distance and varies from 64% for crossing at 500 m to about 83% for crossing at 1000 m or 2000 m.

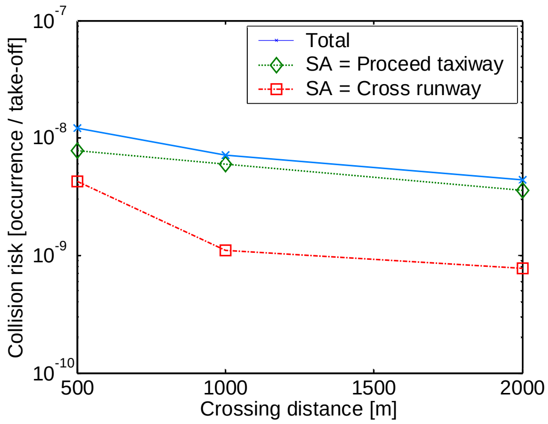


Figure 3. *Contributions to the total collision risk by the simulation model for the cases that the SA of the PF of the taxiing aircraft is to proceed on a taxiway, or to cross the runway.*

Table 2. *SA Dependent collision risk by the simulation model for crossing at a distance of 1000 m (event condition is not distant dependent).*

Probability per take-off	SA by PF of taxiing aircraft	
	Proceed taxiway	Cross runway
Probability of event	$3.5\ 10^{-5}$	$2.0\ 10^{-4}$
Conditional coll. risk	$1.7\ 10^{-4}$	$5.5\ 10^{-6}$
Collision risk	$6.0\ 10^{-9}$	$1.1\ 10^{-9}$

The collision risk in Table 2 depends on the probability of the particular SA condition and the probability of a collision given this condition, for a crossing distance of 1000 metres. The probability of the situation that a pilot taxies across the stopbar not knowing he is approaching the runway, is assumed to be a factor 5.7 smaller than the probability of the situation that the pilot starts crossing the runway while not allowed to do so. Nevertheless, the largely enhanced conditional collision risk leads to a larger collision risk in the former case. The reduced conditional collision risk in the latter situation is due to better monitoring process of the pilot

flying of the taxiing aircraft, if its crew is aware to be heading towards a crossing of an active runway.

Based on results of the accident risk model, it is possible to attain insight in the accident risk reducing performance of involved human operators and technical systems. Table 3 shows conditional collision risks for the situation that an aircraft taxies towards a runway crossing at a distance of 1000 m from the runway threshold while the pilot is aware to taxi on a normal taxiway. The conditional collision risks in Table 3 refer to cases in which the involved human operators either do ('yes') or do not ('no') actively monitor for traffic conflicts. A risk reduction percentage is determined by comparing the conditional collision risk with the situation in which none of the human operators is actively monitoring. In this case, a collision is only avoided by the lucky circumstances that the taxiing aircraft just passes in front of or behind the taking-off aircraft (case 0 in Table 3).

Table 3. *Risk reduction achieved in the simulation model by various combinations of involved human operators for the situation that the pilot flying of the taxiing aircraft intends to proceed on a normal taxiway. See main text for further explanation.*

Case	PF taxiing aircraft	PF taking-off aircraft	Runway controller	Conditional collision risk	Risk reduction
0	no	no	no	$8.9\ 10^{-2}$	-
ATC alert systems on					
1	yes	yes	yes	$1.7\ 10^{-4}$	99.8%
2	yes	no	yes	$4.0\ 10^{-4}$	99.6%
3	no	yes	yes	$9.4\ 10^{-3}$	89.4%
4	yes	yes	no	$2.3\ 10^{-4}$	99.7%
ATC alert systems down					
5	yes	yes	yes	$2.2\ 10^{-4}$	99.8%
6	yes	no	yes	$1.7\ 10^{-3}$	98.1%
7	no	yes	yes	$1.1\ 10^{-2}$	87.9%
8	yes	yes	no	$2.3\ 10^{-4}$	99.7%

A number of model-based insights can be attained by comparing the results of Table 3.

- It follows from case 1 that 99.8% of the accidents can be prevented by the combined effort of all human operators and alert systems.
- It follows from a comparison of cases 1 and 5 that in the normal situation that all human operators are actively monitoring, ATC alert systems

(runway incursion or stopbar violation) almost have no effect on the achieved risk.

- It follows from a comparison of cases 1 and 4, and cases 5 and 8, that the risk reduction that can be achieved by the tower controller in addition to the risk reduction of both pilots is very small.
- It follows from comparison of cases 1 and 3, and cases 5 and 7 that the pilot of the taxiing aircraft has the largest capability to prevent a collision in this context.

Discussion

The accident risk assessment methodology and the associated human performance modelling approach that are discussed in this paper, provide a systematic approach to risk assessment of operations with multiple, dynamically interacting agents. The combined effect of dynamically interacting agents is hard to assess by static or single-agent approaches. As an example, during an argumentation-based risk assessment of the discussed active runway crossing operation, pilots and controllers were asked to estimate their potential to prevent a collision as result of a runway incursion. Especially the contribution of the tower controller was overestimated, because this expert-based evaluation had difficulty to account well for the timing of actions of the pilots and controller. Through Monte Carlo simulations it has become clear that in good visibility conditions, a large part of conflicts is recognised and handled by the pilots before the controller can react.

By definition a model is unequal to reality. Hence, application in a risk assessment of the discussed models requires an evaluation of the effect on the risk of the assumptions adopted in the modelling process (Everdij and Blom, 2002). This evaluation takes into account the particular context of the operation assessed and will be conducted in a follow-up study. Then interviews with pilots and controllers will be conducted to obtain their feedback on the assumptions made. In these interviews, typically asked questions will refer to single-agent tasks and aspects such as task duration. These kind of questions can be more easily estimated than small probability values (e.g., conflict resolution probability estimates in a multi-agent environment) such as typically included in interviews for argumentation-based risk assessment.

The feasibility of using human performance modelling in accident risk assessment for a conflict scenario with a considerable number of interacting humans and technical systems has been illustrated for an active runway crossing operation. The model results stress the importance of proper situation awareness of the pilots flying for minimising runway incursion risk.

References

Amalberti R, Wioland L (1997). Human error in aviation. In Soekkha H (ed.), Aviation Safety, pp. 91-108.

Blom HAP, Bakker GJ, Blanker PJG, Daams J, Everdij MHC, Klompstra MB (2001). Accident risk assessment for advanced air traffic management. In: Donohue GL and Zellweger AG (eds.), Air Transportation Systems Engineering, AIAA, pp. 463-480.

Blom HAP, Stroeve SH, Everdij MHC, Van der Park MNJ (2003). Human cognition performance model to evaluate safe spacing in air traffic, Human Factors and Aerospace Safety, Vol. 3, pp. 59-82.

Blom HAP, Stroeve SH (2004). Multi-agent situation awareness error evolution in air traffic. Proc. Probabilistic Safety Assessment and Management, Berlin, Germany

Cacciabue PC (1998). Modelling and simulation of human behaviour in system control. Springer-Verlag, Berlin.

Corker KC, Blom HAP, Stroeve SH (2005). Integration of Air-MIDAS and TOPAZ human performance modelling for air traffic accident risk assessment. Proc. 13th International Symposium on Aviation Psychology, Oklahoma City, USA.

Endsley MR (1995). Towards a theory of situation awareness in dynamic systems. Human Factors, 37(1), 32-64.

Everdij MHC, Blom HAP. Bias and uncertainty in accident risk assessment. National Aerospace Laboratory NLR, Report NLR-TR-2002-137, 2002.

Hollnagel E (1993). Human reliability analysis, context and control. Academic Press, London.

Stroeve SH, Blom HAP, Van der Park MNJ (2003). Multi-agent situation awareness error evolution in accident risk modelling. 5th USA/Europe ATM R&D Seminar, Budapest.

Wickens CD, Mavor AS, Parasuraman R, McGee JP (1998). The future of Air Traffic Control, Human Operators and Automation. National Academy Press, Washington DC.

Wickens CD, Hollands JG (1999). Engineering psychology and human performance. Prentice Hall.

THE EFFECTIVENESS OF A PERSONAL COMPUTER AVIATION TRAINING DEVICE (PCATD), A FLIGHT TRAINING DEVICE (FTD), AND AN AIRPLANE IN CONDUCTING INSTRUMENT PROFICIENCY CHECKS

Henry L. Taylor, **Donald A. Talleur**, **Esa M. Rantanen**, and **Tom W. Emanuel, Jr.**
University of Illinois at Urbana-Champaign
Institute of Aviation

This project evaluated the effectiveness of a personal computer aviation training device (PCATD), a flight training device (FTD) and an airplane for conducting an instrument proficiency check (IPC). The study compared the performance of pilots receiving an IPC in a PCATD, in a FTD and in an airplane (IPC #1) with performance on a later IPC in an airplane (IPC #2). Chi-square tests were used to analyze the IPC #1 and IPC #2 data to determine whether the treatment (assignment to group) had an effect on the pass/fail ratio for the IPC #1 and IPC #2 flights respectively. The treatment effect on the IPC #1 and IPC #2 pass/fail ratios were not statistically significant. A series of planned-comparison tests were performed both between the experimental groups and between subjects within each experimental group. The PCATD group was compared to the Airplane group and to the FTD group, the Airplane group to the FTD group. None of these comparisons showed statistically significant ($a < .05$) differences between groups. These findings provide compelling evidence for permitting the use of PCATDs to give IPCs.

Introduction

To maintain instrument currency, instrument pilots must meet the recency of experience requirements of FAR 61.57(c) or (d) every six months. The recency of experience requirements may be conducted in an airplane or simulated in an approved flight training device (FTD). If an instrument pilot fails to meet recency of experience requirements within a 12-month period, an instrument proficiency check (IPC) must be accomplished with a certified flight instructor, instrument (CFII) to regain instrument currency.

Taylor et al. (1996, 1999) conducted a study to determine the extent to which a personal computer aviation training device (PCATD) could be used to develop specific instrument skills that are taught in instrument flight training and to determine the transfer of these skills to the aircraft. This research led to an additional study by the Institute of Aviation of the University of Illinois at Urbana-Champaign (UIUC) to determine the effectiveness of PCATDs for maintaining instrument currency (Taylor et al., 2001; Talleur, Taylor, Emanuel, Rantanen, and Bradshaw, 2003). In the latter study, a total of 106 instrument current pilots were divided in four groups. The pilots in each group received an instrument proficiency check (IPC #1). During a six-month period following IPC #1, the pilots in three groups received recurrent training in a PCATD, a Frasca flight training device (FTD), or an airplane, respectively. The fourth (control) group received no training during the six-month period. After this time, the pilots in each group flew an instrument proficiency check (IPC #2). The comparison of IPC #1 and IPC #2 indicated that both the PCATD and the Frasca FTD were more effective in maintaining instrument proficiency when compared to the control group and at least as effective as the airplane. The study also found that of 106 instrument current pilots, only 45 (42.5%) were able to pass IPC #1. Of the group who received an IPC in a Frasca FTD to regain currency, only 22 of 59 were able to subsequently able to pass IPC #1 in an airplane. This study established the effectiveness of PCATDs for use in instrument currency training. However, the question of whether PCATDs are effective for administering the IPC has not been demonstrated. Based on the data above, a question concerning the effectiveness of the Frasca FTD in administrating an IPC also arises.

The purpose of the present study was to compare the performance of pilots receiving an IPC in a PCATD, a FTD or an airplane (IPC #1) with their performance in an airplane (IPC #2). The comparison of performance in a PCATD to that in an airplane investigated the effectiveness of the PCATD as a device in which to administer an IPC. Currently, the PCATD is not approved to administer IPCs. The comparison of performance in a FTD with performance in an airplane will helped determine whether the current rule to permit IPCs in a FTD is warranted. Finally, the comparison of performance of pilots receiving IPC #1 in an airplane and IPC #2 in an airplane with a second CFII permitted the determination of the reliability of IPCs conducted in an airplane.

Method

Subjects

Seventy-five pilots participated in the study (25 subjects in each group; FTD, PCATD and airplane). Most of the participating pilots were instrument current but a few fall into one of three other categories of instrument currency: (1) within one year of currency, (2) outside of one year of currency but within two years of currency, and (3) outside two years but within five years of currency.

A limited number of pilots who were more than two years out currency received an average of six hours training equally distributed among the FTD, PCATD and airplane to prepare them for the IPC. This procedure was discontinued after the second year to reduce expenses, and no additional subjects of this currency status were added to the project.

Equipment

Two FAA-approved Elite PCATDs and one FAA-approved Frasca 141 FTD with a generic single-engine, fixed gear, fixed-pitch propeller performance model were used in the study. The FTD is approved for instrument training towards the instrument rating, instrument recency of experience training, and IPCs as well as for administering part of the instrument rating flight test. Two single-engine 180 hp Beechcraft Sundowner aircraft (BE-C23) with fixed-pitch propellers and fixed undercarriage were used as the aircraft for IPC #1 and IPC #2.

Procedure

All participants received a familiarization flight and a review of the systems and instrumentation in the FTD, the PCATD and the airplane prior to being assigned to an experimental group. Following the familiarization flights, subjects were assigned to one of the three groups (FTD, PCATD and Airplane) with a constraint that the currency categories were balanced among the groups. All 75 pilots received a baseline IPC flight in the FTD, PCATD or an airplane (IPC #1) according to their group assignment. Table 1 depicts the experimental design.

The IPC is a standardized test of the instrument pilot's instrument skills. The types of maneuvers, as well as completion standards for an IPC, are listed in the instrument rating practical test standards (PTS) (U.S. Department of Transportation, 1998). A flight scenario that follows the current guidelines (at that time) for the flight maneuvers required by the PTS was used for the IPC. This scenario was used to collect baseline data and to establish the initial level of proficiency for each subject who participants in the project.

The IPC flights contained six maneuvers (VOR approach, holding pattern, steep turns, unusual altitude recovery, ILS approach and a partial-panel non-precision approach). ATC communication procedures are also scored. The CFIIs for the IPC #1 flight used a form that was designed to facilitate the collection of three types of data (Phillips et al., 1995). First, within each maneuver there were up to 24 variables (e.g., altitude, airspeed) that were scored as pass/fail indicating whether performance on those variables met PTS requirements. Second, the flight instructor judged whether the overall performance of the each maneuver was pass/fail. Third, the CFII recorded if the overall performance of the subject met the PTS for the IPC. The instructors who administered the IPC #1 flight were standardized on the scenario to be flown and the scoring procedure.

IPC #1 was flown with a certified flight instructor, instrument (CFII) who acted both as a flight instructor and as an experimental observer. The participants are required to refrain from instrument flight following IPC #1 until IPC #2 is completed. They must also agree not to use a PCATD or a FTD for instrument training during this period.

Table 1. *Experimental Design*

Group	Fam. Flight	Initial IPC (IPC#1)	Final IPC (IPC#2)
Airplane	In Airplane In Frasca In Elite	IPC flight in Sundowner	IPC flight in Sundowner
Frasca	In Airplane In Frasca In Elite	IPC flight in Frasca	IPC flight in Sundowner
PCATD	In Airplane In Frasca In Elite	IPC flight in Elite	IPC flight in Sundowner

After a period not exceeding two weeks, all subjects flew a final IPC (IPC #2) in the aircraft to assess instrument proficiency. IPC #2 was conducted by a different CFII than IPC #1 to eliminate experimenter bias. The CFII for IPC #2 was blind to both the group to which the subject belonged and to the subject's performance on IPC #1. In terms of maneuvers, IPC #2 was identical to IPC #1. This final session contained all required maneuvers that a pilot must satisfactorily complete in order to receive an endorsement of instrument proficiency. Completion of IPC #2 marked the end of a subject's involvement in the experiment.

Results

The pass/ fail rates by group for the 75 subjects for IPC #1 and IPC #2 are shown in Table 2, presenting the number and percentage of pilots that passed/failed IPC #1 and IPC #2 for each of the three experimental groups and for the total subjects.

Table 2. *Pass/Fail rates by group*

IPC#1					
Group	N	Pass	(%)	Fail	(%)
Aircraft	25	6	(24)	19	(76)
FTD	25	9	(36)	16	(64)
PCATD	25	9	(36)	16	(62)
Total	75	24	(32)	51	(68)

IPC#2

Group	N	Pass	(%)	Fail	(%)
Aircraft	25	13	(52)	12	(48)
FTD	25	14	(56)	11	(44)
PCATD	25	15	(60)	10	(40)
Total	75	42	(56)	33	(44)

Figures 1 and 2 show the differences between pass rates for the three groups for IPC #1 and IPC #2, respectively. Inspection of Figures 1 and 2 indicate few differences between groups for the number of participants who passed IPC #1 and IPC #2. A total of 24 of 75 subjects (32%) passed the IPC #1 flight in the airplane, FTD and PCATD and a total of 42 of 75 subjects (56%) passed the IPC #2 flight.

Chi-square tests were used to analyze the IPC #1 and IPC #2 data to determine whether the treatment (assignment to group) had an effect on the pass/fail ratio for the IPC #1 and IPC #2 flights respectively. The treatment effect on the IPC #1 pass/fail ratios, c2 (2, N=75) = 0.32, p = 0.85, and on IPC #2 pass/fail ratios, c2 (2, N=75) = 1.1, p = 0.58 were not statistically significant. A series of planned-comparison tests were performed between and among the experimental groups but one showed significant differences between the groups (p > .10).

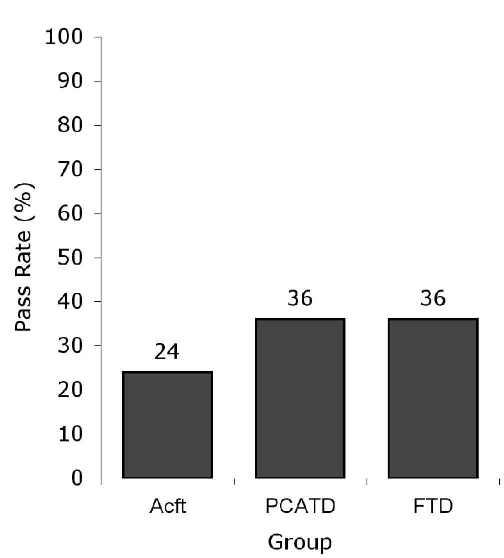


Figure 1. *Pass rates in IPC #1 for the experimental groups*

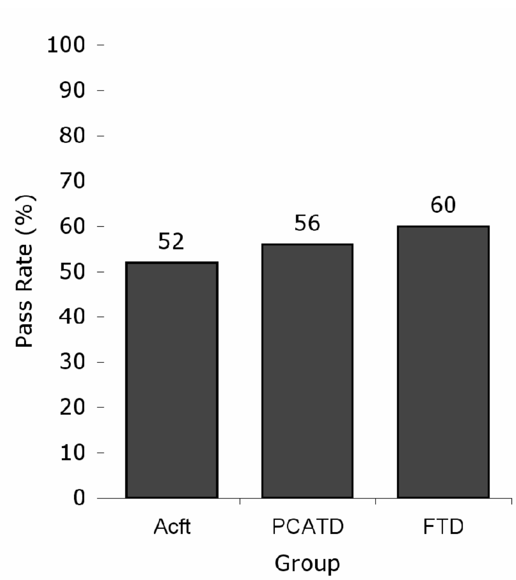


Figure 2. *Pass rates in IPC #2 for the experimental groups.*

The pass/fail rates by currency status are shown in Table 3. A total of 53 current pilots took IPC #1 and 19 passed (36%) while 34 failed (64%). Of the 53 current pilots taking IPC #2 and 30 passed (57%) while 23 failed (43%).

Analysis of the change of performance that took place between the IPC #1 and IPC #2 flights was made in order to understand the effectiveness of the three devices for conducting IPCs. Chi-square tests for changes in performance between IPC #1 and IPC#2 were used to determine if subjects' performance had improved or deteriorated between the two sessions. All three experimental groups showed no significant changes in performance between IPC #1 and IPC #2, ($p > .05$).

Table 3. *Pass/Fail rates by currency*

IPC #1					
Currency	N	Pass	(%)	Fail	(%)
Current	53	19	(36)	34	(64)
Within 1 year	7	2	(29)	5	(71)
Within 1-2 years	1	1	(100)	0	(0)
2-5 years	14	2	(14)	12	(86)

IPC #2					
Currency	N	Pass	%)	Fail	(%)
Current	53	30	(57)	23	(43)
Within 1 year	7	6	(86)	1	(14)
Within 1-2 years	1	1	(100)	0	(0)
2-5 years	14	5	(36)	9	(64)

It was expected that performance on IPC #1 would be a good predictor of performance on IPC#2. Table 4 shows a comparison of the pass/fail rates for IPC #1 and IPC #2. Of the 24 participants who passed IPC #1 only 14 also passed IPC #2 (58%), and of the 51 participants who failed IPC #1 only 23 (45%) subsequently failed IPC #2 (a total of 37). Twenty-eight participants, who failed IPC #1 subsequently passed IPC #2 and 10 of the participants who passed IPC #1 subsequently, failed IPC #2 (a total of 38). Therefore, performance on IPC #1 predicted the performance on IPC# 2 only at the chance level. Indeed, the McNemar change in performance analysis between IPC #1 and IPC #2 for all participants was significant; c2 (1, N = 75) = 8.53, p = .004.

Table 4. *IPC #1 vs. IPC #2 Pass/Fail*

		IPC#2		
		Pass	Fail	Total
IPC#1	Pass	14	10	24
	Fail	28	23	51
	Total	42	33	75

Discussion

Reliability of FTDs and PCATDs for IPC

This study revealed no significant differences in performance by instrument pilots on an IPC given in either a PCATD, and FTD or an airplane. No significant difference was found on IPC #1 among the three groups, which indicates that the participants were likely to pass or fail an IPC in an Airplane as often as either the PCATD or the FTD. In addition there was no significant difference on IPC #2 indicating that the device in which the participants had IPC #1 had no influence on their pass/fail rates on IPC #2 in the airplane. The planned comparisons showed that pass/fail rates on IPC #2 of the PCATD group was statistically indistinguishable from both the airplane and the FTD groups. In addition, there was no difference in pass/fail rates between the aircraft and the FTD groups. These findings present compelling evidence for permitting the use of PCATDs to give IPCs.

Pre-Test—Post-Test Reliability

It was expected that performance on IPC #1 would be a good predictor of performance on IPC#2. However, a comparison of the pass/fail rates for IPC #1 and IPC #2 indicated that the performance on the baseline IPC did not predict performance on the final IPC. Only 58 percent of the participants who passed IPC #1 also passed IPC #2 and only 45 percent of the participants who failed IPC #1 also failed IPC #2. Only 49 percent of the participants either passed both tests or failed both tests, while 51 percent of the participants passed IPC #1 and failed IPC #2 or failed IPC #1 and passed IPC #2. Therefore performance on IPC #1 predicts performance on a second IPC at a chance level.The McNemar change in performance between IPC #1 and IPC #2 for all participants was significant but the comparisons for the individual three groups were not significant. Some of the failures may be related to a lack of familiarity with the PCATD, the FTD and the Sundowner airplane, since few of the participants had flown either of the devices prior to the study. The familiarization flights in each of the devices were expected to provide sufficient familiarity with the devices to eliminate the problem but may have failed to do so. It is possible that additional familiarity with instrument flying in each device, in addition to the VFR familiarization, was needed. The former was not done in order to minimize a possible training effect on group assignment.

Instrument Currency vs. Instrument Proficiency

Of the 53 participants who were instrument current, only 19 (36 %) passed IPC #1. The earlier study by Taylor et al. (2001) and Talleur et al. (2003) showed that 42 % of the instrument current pilots passed the initial IPC. The results from the current study are only slightly worse in this regard than those from earlier studies. In addition, most of the participants tested in the previous study had not taken an IPC after the test was standardized to include required maneuvers (thereby increasing the difficulty of the IPC test). This finding raises questions concerning the relationship between instrument currency and instrument proficiency. Less than half of the participants were able to demonstrate instrument proficiency in an IPC in the airplane. This suggests the need for the FAA to consider changing the recency of experience requirements for instrument currency. Taylor et al., (2001) made the same observation and the current study reinforces the concern that currency rules are inadequate for instrument pilots to maintain proficiency. As Taylor et al., (2001) suggested, an alternative approach would be to require a periodic IPC to demonstrate instrument proficiency in addition to the current currency requirements.

Acknowledgments

This work is supported under Federal Aviation Administration (FAA) Cooperative Agreement 2001-G-037. The study was sponsored by FAA Headquarters Flight Standards Service, General Aviation and Commercial Division. Dennis Beringer serves as the COTR. Views expressed herein do not necessarily represent official FAA positions. We express our appreciation Ms. Mary Wilson who scheduled subjects and to Catherine Trock and Joani Disilvestro for their assistance with the data. We also thank the Institute of Aviation flight instructors who provided instrument training in the FTD, PCATD and the aircraft, the Institute flight instructors who served as IPC check pilots, and the instrument pilots for their participation in the study.

References

Lendrum, L., Taylor, H. L., Talleur, D. A., Hulin, C. L., Bradshaw, G. L., & Emanuel, T. W. (2000). *IPC Data Logger Operation Manual* (ARL-00-8/FAA-00-5). Savoy, IL. University of Illinois at Urbana-Champaign

Talleur, D. A., Taylor, H. L., Emanuel, T. W., Jr., Rantanen, E, M., & Bradshaw, G. L. (2003). Personal computer aviation training devices: Their effectiveness for maintaining instrument currency. *The International Journal of Aviation Psychology*, *13*(4), 387-399.

Taylor, H. L., Talleur, D. A., Bradshaw, G. L., Emanuel, T. W., Rantanen, E. M., Hulin, C. L. & Lendrum, L. (2001). *Effectiveness of personal computers to meet recency of experience requirements*. (Technical Report ARL-01-6/FAA-01-1). Savoy, IL. University of Illinois, Aviation Research Laboratory

Taylor, H. L., Lintern, G., Hulin, C. L., Talleur, D. A., Emanuel, T. W. & Phillips, S. I. (1996). *Transfer of training effectiveness of personal computer-based aviation training devices* (Technical Report ARL-96-3/FAA-96-2). Savoy, IL. University of Illinois, Aviation Research Laboratory.

Taylor, H. L., Lintern, G., Hulin, C. L., Talleur, D. A., Emanuel, T. W., Jr., & Phillips, S .I. (1999) Transfer of training effectiveness of a personal computer aviation training device. *The International Journal of Aviation Psychology*, *9*(4), 319-335.

U.S. Department of Transportation. (1997). *Qualification and approval of personal computer-based aviation training devices, AFS-840* (Advisory Circular No: AC 61-126). Washington, DC: Federal Aviation Administration, U.S. Department of Transportation.

U.S. Department of Transportation. (1998, October). *Instrument rating for airplane, helicopter and airship: Practical test standards* (FAA-S-8081-4C). Washington, DC: Federal Aviation Administration, U.S. Department of Transportation.

TRANSFER OF TRAINING EFFECTIVENESS OF A FLIGHT TRAINING DEVICE (FTD)

Henry L. Taylor, **Donald A. Talleur**, **Tom W. Emanuel, Jr.**, and **Esa M. Rantanen**
Institute of Aviation, University of Illinois at Urbana-Champaign
Savoy, Illinois

A transfer of training research design was used to measure the effectiveness of a flight training device (FTD) and to determine the point at which additional training in a FTD was no longer effective. The dependent measures were number of trials to specific completion standards, time to complete a flight lesson, and time to a successful evaluation flight. Percent transfer and transfer effectiveness ratios (TERs) were computed for each instrument task and for the time to complete a flight lesson. The data from the current study indicates that the FTD and the PCATD appear effective in teaching basic and advanced instrument tasks to private pilots but the limited number of subjects prevented this effectiveness from being convincingly demonstrated. As a result of prior training in an FTD and a PCATD time to a stage check or an instrument rating flight check flight was less when compared to an airplane control group.

Introduction

In an earlier study by Taylor et al., (1996), a commercially available Personal Computer Aviation Training Device (PCATD) was evaluated in a transfer of training experiment to determine its effectiveness for teaching instrument tasks. The data indicated that transfer savings for both the number of trials to reach a performance criterion for instrument tasks and time to complete a flight lesson were positive and substantial for new instrument tasks. A comparison of instrument rating course completion times resulted in a saving of about four hours in the airplane as a result of prior training in the PCATD. As a result of the Taylor et al. (1996) study, a Federal Aviation Administration (FAA) advisory circular published in 1997 permits 10 hours of instrument training to be completed in an approved PCATD.

To evaluate transfer of training effectiveness of a flight training device (FTD), the performance of subjects trained on instrument tasks in an FTD and later trained to criterion in an airplane must be compared to the performance of subjects trained to criterion only in the airplane. Roscoe (1971) demonstrated that the transfer effectiveness ratio (TER) accounts for the amount of prior training in ground trainers by specifying the trials/time saved in the airplane as a function of the prior trials/time in the ground training. Because diminishing transfer effectiveness ratios as the number of trials or hours in ground trainer increases, additional ground-based training will at some point cease to be cost effective. The law of diminishing returns adequately describes this relationship between extra training and resultant benefit. The purpose of the present study was to use an incremental transfer of training research design to measure the effectiveness of an FTD and a PCATD to determine the point at which additional training in a FTD or a PCATD is no longer effective.

Method

Participants

Participants were assigned to four FTD (Frasca) groups, one PCATD group, and a control (airplane) group. In the initial proposal a total of 180 pilots (30 in each of the 6 groups) were scheduled to participate in the study. Due to funding reductions in the second and third years, the number of pilots in the study was first reduced to a total of 120 pilots (20 subjects in each group) and due to the elimination of FY 2005 funding the eventual number of participants for each group who successfully completed the instrument program ranged between 15 and 20. The participants were University of Illinois, Institute of Aviation private pilot students, who were enrolled in the Institute's instrument flight program. This program consists of two semester courses: AVI 130, Basic Instruments and AVI140, Advanced Instruments. All students in the instrument program were involved in the study. A total of 106 students completed the study. Each semester the students were assigned equally to the six groups while maintaining a balanced number of subjects across all groups to account for students who did not complete the course prior to completion.

Equipment

Training in the FTD was conducted in four Frasca 141 FTDs with generic single-engine, fixed-gear, and fixed-pitch propeller performance models. The PCATD training was conducted using FAA approved PCATDs from Aviation Teachware Technologies (ELITE) v. 6.0.2, with flight controls by Precision Flight Controls. These PCATDs simulated the flight characteristics of the Piper Archer III aircraft. Airplane training was carried out in the Piper Archer III aircraft, which is a single-engine, fixed-pitch propeller, fixed undercarriage aircraft.

Procedure

The Frasca groups received 5, 10, 15, and 20 hours of prior instrument training in a FTD, respectively, and the PACTD group received 5 hours of prior training in the ELITE PCATD. .With the exception of the cross country training for Frasca groups 15 and 20 the prior training was distributed equally between AVI 130 and AVI 140. A Control group received all training in the airplane. Training on selected instrument tasks using the FTD and PCATD was administered to the four FTD groups and the PCATD group during four flight lessons for each semester. In addition, FTD training was given during certain x-country lessons in both AVI 130 and AVI 140 for the 15 and 20 hour FTD groups.

Prior to the start of each semester, all flight instructors were standardized on the use of the FTD and PCATD, changes in the training course outlines (TCOs), and experimental procedures. Flight instructors served as both instructors and data collectors. They rated student performances on designated flight tasks in the aircraft. For performance assessment in the aircraft, each instructor recorded if the student met the completion standards during the execution of the designated flight tasks. They also recorded the number of trials to criterion for specific tasks and flight time to complete a flight lesson (Phillips et al., 1995). Four check pilots, blind to the allocation of students to training conditions, were used to conduct the AVI 130 stage check and the AVI 140 instrument rating flight check.

Each flight instructor was instructed to schedule a stage check after Flight Lesson 40 in AVI 130, and an instrument rating flight check after Flight Lesson 55 in AVI 140 when the student was judged to be able to meet the proficiency standards for the stage check and the instrument proficiency check, respectively. These check flights permitted the assessment of the differential time to complete the flight course as a function of the amount of prior training in the FTD and the PCATD. Those students who failed the evaluation flight or failed to meet the proficiency standards by Flight Lesson 45 (stage check) and Flight Lesson 60 (instrument rating check flight) were provided additional flight time to reach proficiency. Dependent measures were trials in the airplane to proficiency, time to complete the flight lessons in the airplane, and total course completion time in the airplane for both courses.

Mean number of trials to reach criterion in the airplane for selected instrument tasks, and mean time to complete the flight lesson in the airplane were computed for all groups for both courses. Analyses of Variance (ANOVA) were performed to analyze the differences between the six groups. ANOVA were used to determine the significance of the trial variable and flight lesson completion time variable as a function of experimental treatment for both AVI 130 and AVI 140. Finally, ANOVA were used to determine the significance of the differences of the time to a successful check flight for the AVI 130 and AVI 140 courses as a function of the experimental treatment for the three groups (PCATD, FTD 5 and 10 groups) that received only prior training only on instrument tasks compared to the control group. To further identify the locus of any significant effects, post–hoc tests were used to make specific pairwise comparisons using Tukey's test of significance.

Results

A total of 124 subjects successfully completed the AVI 130 Basic Instruments course and took the final check ride. Table 1 shows the results of the check ride for the six groups. A total of 75 students passed the check ride on the first attempt and 49 students passed on the second attempt. Nine students were recommended for a remedial course, AVI 102. The total dual flight time to completion for the six groups is shown in Table 1 and in Figure 1. The average dual flight time to course completion for the airplane group was greater than the average time for each of the five experimental groups who had prior training in the PCATD or the FTD. The airplane group required 22.35 hours of dual to complete the course while the five experimental groups, after prior training in the PCATD or the FTD, required between 18.31and 20.87 hours of dual flight time in the airplane to complete the course.

For AVI 130, ANOVAs were computed to determine effect of the experimental treatment (assignment to groups) for mean trials to criterion in the airplane for selected instrument tasks for the four flight lessons for the three groups (PCATD, FTD 5 and 10 groups), that received prior training only on instrument tasks, and the control group. For Flight Lesson 37, there was a significant difference for both ILS and VOR ($F(3,81)=2.78$; $p < .05$ and $F(3,81)=5.12$; $p < .05$ respectively) and for Flight Lesson 38 there was a significant difference for VOR and DME ARC ($F(3,81)=2.84$; $p < .05$ and $F(3,81)=2.70$; $p < .05$ respectively). No other instrument tasks were significant. For Flight Lesson 37, pairwise comparisons using Tukey's test of significance indicated a significant difference between the airplane and the Frasca 5 and 10 groups ($p < .05$). ANOVA were computed to determine effect of the experimental treatment for mean time to complete the flight lesson for the four flight lessons for the PCATD, FTD 5 and 10 groups and the control group. A significant treatment effect was found for Flight Lessons 34/35, 36, and 37 (all $p < .05$). Pairwise comparisons indicated a significant difference between the airplane and all three groups for Flight Lesson 34/35 and between the Airplane and the Frasca 5 and 10 groups for Flight Lesson 37 (both $p < .05$). An ANOVA to determine effect of the experimental treatment for total course completion time in the airplane was computed. A significance difference was found (F (3,80)=3.67; p <.05. Pairwise comparisons using indicated a significant difference between the airplane and the Frasca 5 group ($p < .05$).

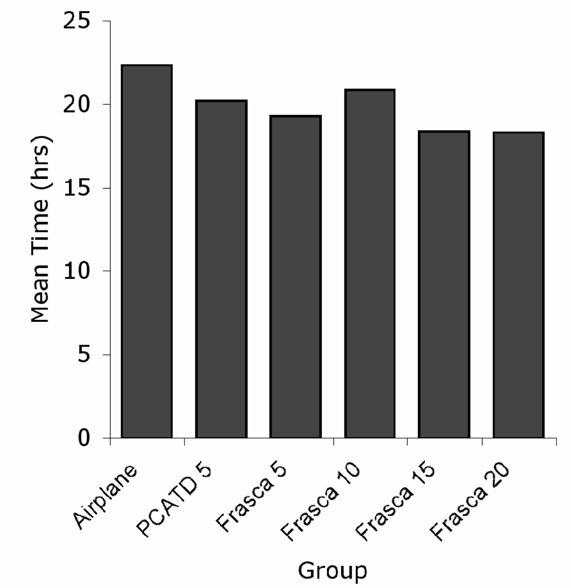


Figure 1. *Total time to successful completion of flight lesson 45, showing incremental transfer effectiveness of the experimental groups.*

A total of 106 subjects successfully completed the AVI 140, Advanced Instruments course and took the final check ride (the instrument rating flight check). Table 2 shows the results of the check ride. A total of 51 students passed the check ride on the first attempt and 46 students passed on the second attempt. The total dual flight time to completion for the six groups for the advance instrument course (AVI 140) is shown in Table 2 and in Figure 2. The average course completion time for the airplane group is greater for each of the five experimental groups who had prior training in the PCATD or the FTD. The airplane group required 26.38 hours of dual to complete the course while the total dual hours in the airplane to completion for the five experimental groups ranged from 25.78 to 20.79 hours after prior training in the PCATD or the FTD.

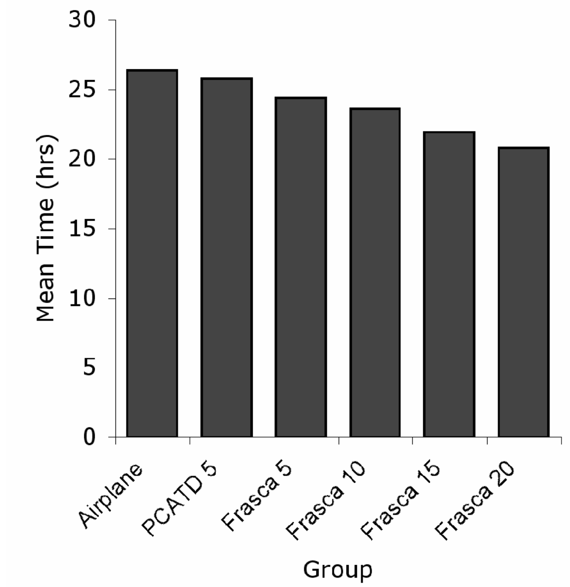


Figure 2. *Total time to successful completion of flight lesson 60, showing incremental transfer effectiveness of the experimental groups.*

For AVI 140, ANOVAs were computed to determine effect of the experimental treatment (assignment to groups) for mean trials to criterion in the airplane for selected instrument tasks for the four flight lessons for the three groups (PCATD, FTD 5 and 10 groups), that received prior training only on instrument tasks, and the control group. For Flight Lesson 48, there was a significant difference for ILS approach (F (3.77)=2.90; $p < .05$). Pairwise comparisons indicated a significant difference between the PCATD 5 and the Frasca 5 group ($p < .05$). For Flight Lesson 50, there was a significant difference for NDB approach (F (3,77)=3.90; $p < .05$). Pairwise comparisons indicated a significant difference between the Airplane and the PCATD 5 and the Frasca 5 groups ($p < .05$). For Flight Lesson 52, there was a significant difference for NDB Hold and GPS approach (F (3,76)=3.34; $p < .05$ and F (3,75)=3.14; $p < .05$ respectively). Pairwise comparisons indicated a significant difference between the PCATD 5 and the Frasca 5 groups for NDB Hold ($p < .05$). ANOVAa were computed to determine effect of the experimental treatment for mean time to complete the flight lesson for the four scored flight lessons for each of the three groups (PCATD, FTD 5 and 10 groups) that received only prior training on instrument tasks and the Control group. A significant treatment effect was found for Flight Lesson 52 (F (3,76)=5.79; $p < .05$). Pairwise comparisons indicated a significant difference between the PCATD 5 and the Frasca 5 and 10 groups ($p < .05$). An ANOVA was computed to determine effect of the experimental treatment for total course completion time in the airplane for AVI 140. A significance difference was found ($F(3,65)=2.77$; $p < .05$). Pairwise comparisons indicated no significant difference between any groups.

The effect of allocating 5 and 10 hours in the Frasca for cross-country flight was evaluated. For AVI 140, the airplane group required 26.38 hours of dual to completion while the Frasca 10,15 and 20 groups required 23.60, 21.93 and 20.79 hours respectively. This represents a savings of 2.78 hours, 4.45 hours and 5.59 hours respectively. Since the Frasca 15 and 20 groups received the same treatment as the Frasca 10 group regarding training only on instrument tasks and an additional 5 and 10 hours respectively for cross country training, the computed savings for the 5 and 10 hours cross country time was 1.67 and 2.81 hours respectively.

Discussion

The data from the current study indicates that the FTD and the PCATD appear effective in teaching basic and advanced instrument tasks to private pilots but the limited number of subjects prevented this effectiveness from being convincingly demonstrated. With the limited number of subjects and the current variability among subjects, the power of the ANOVA is low. The current data fail to replicate the findings of Taylor et al. (1996, 1999) that PCATDs are useful to teach instrument tasks to private pilots. As a result of prior training in an FTD and a PCATD, time to the stage check in AVI 130 and to the instrument rating flight check was less for three groups (PCATD, FTD 5 and 10 groups) that received prior training only on instrument tasks as compared to the control group. For AVI 130, pairwise comparisons indicated a significant difference between the airplane and the Frasca 5 group and for AVI 140, pairwise comparisons indicated no significant difference between any groups. One purpose for conducting an incremental transfer of training study is to determine at what point additional training in the FTD and the PCATD is no longer effective. The data collect does not permit this to be determined convincingly. A study by Taylor et al., (2002) clearly indicated that the use of 5 hours of PCATD time was cost-effective based on the allocation of PCATD time for these tasks for the PCATD 5 group. The current study shows that the PCATD is only effective for the NDB task. We attribute the difference between the two studies to be the result of the lack of power in the current study.

Time to complete the flight lesson was significant for three flight lessons out of four for AVI 130 when comparing the PCATD, FRASCA 5 and 10 groups with the Control group, but for only one flight lesson out of four for AVI 140. Taylor, et al (2002), which tested the incremental effectiveness of the PCATD, found two of four flight lessons significant for AVI 130 and one for AVI 140.

We do not believe that data generated in the current study provides convincing evidence for flight schools to use in determining how to best implement PCATDs or FTDs in their training programs. There is the possibility that FTDs can be used effectively for teaching cross-country procedures in addition to using them to teach instrument tasks, but the current study has failed to demonstrate significant savings through their use.

Acknowledgments

This work was supported under Federal Aviation Administration (FAA) Cooperative Agreement Number 02-G-033 and sponsored by the FAA Headquarters, Flight Standard Service, General Aviation & Commercial Division. Dennis Beringer served as the COTR. Views expressed herein do not necessarily represent official FAA positions. We express our appreciation to Ms. Sybil Phillips for invaluable assistance with flight operations and with student management. Mr. Bill Jones, Mr. David Boyd, and Don Talleur served as check pilots. We also thank the Institute of Aviation flight instructors and students for their participation in the study.

References

Phillips S. I., Taylor, H. L., Lintern, G., Hulin, C. L., Emanuel, T., & Talleur, D. (1995). Developing performance measures for evaluating personal computer-based aviation training devices within a FAR Part 141 pilot training school. *Proceedings of Aviation Psychology 8th International Symposium*. Columbus, OH: Ohio State U.

Roscoe, S. N. (1971). Incremental transfer effectiveness. *Human Factors*, *13*, 561-567.

Taylor, H. L., Lintern, G., Hulin, C. L., Talleur, D., Emanuel, T., & Phillips, S. (1996). *Transfer of training effectiveness of personal computer-based aviation training devices* (ARL-96-3/FAA-96-2). Savoy, IL: Aviation Research Laboratory.

Taylor, H. L., Lintern, G, Hulin, C. L., Talleur, D. A., Emanuel, T., & Phillips, S. (1999). Transfer of training effectiveness of a personal computer

aviation training device. *International Journal of Aviation Psychology, 9*, 319-335.
Taylor, H. L., Talleur, D. A., Emanuel, T. W., Jr., Rantanen, E. M., Bradshaw, G. L., & Phillips, S. I. (2002). Incremental training effectiveness of personal computers used for instrument training (Final Technical Report ARL-02-5/NASA-02-3). Savoy, IL: University of Illinois, ARL.

Table 1.
Flight Lesson 45 Statistics (Fall, 2002, Spring, Summer, Fall 2003 and Spring 2004)

	Airplane Only	PCATD 5.00	Frasca 5.00	Frasca 10.00	Frasca 15.00	Frasca 20.00
Number of Students	22	20	22	20	21	19
% First Flight Pass Rate	59.00 (N=13)	65.00 (N=13)	45.45 (N=10)	75.00 (N=15)	76.19 (N=16)	42.11 (N=8)
% Second Flight Pass Rate	100.00 (N=9)	100.00 (N=7)	100.00 (N=12)	100.00 (N=5)	80.00 (N=5)	100.00 (N=11)
Students Recommended 102	0	0	1	1	4	3
Total Dual to Completion	22.35 (N=22)	20.20 (N=20)	19.27 (N=22)	20.87 (N=20)	18.36 (N=21)	18.31 (N=19)
Variance Tot. Dual to Completion	9.39	6.40	10.03	14.17	9.87	9.48

Note: This lesson is the final check ride for AVI 130.

Table 2.
Flight Lesson 60 Statistics (Spring, Summer, Fall, 2003, Spring, Summer, Fall 2004)

	Airplane Only	PCATD 5.00	Frasca 5.00	Frasca 10.00	Frasca 15.00	Frasca 20.00
Number of Students	18	18	20	16	15	19
% First Flight Pass Rate	44.44 (N=8)	55.56 (N=10)	45.00 (N=9)	43.75 (N=7)	40.00 (N=6)	57.89 (N=11)
% Second Flight Pass Rate	100.00 (N=10)	75.00 (N=6)	88.89 (N=8)	88.89 (N=8)	100.00 (N=9)	62.50 (N=5)
Students Recommended 102	2	3	4	3	5	2
Total Dual to Completion	26.38 (N=18)	25.78 (N=17)	24.40 (N=18)	23.60 (N=16)	21.93 (N=15)	20.79 (N=18)
Variance Tot, Dual to Completion	16.55	6.03	7.92	8.80	10.20	17.89

AN INTELLIGENT AIRCRAFT LANDING SUPPORT PARADIGM

Steve Thatcher
Aviation Education, Research and Operations Laboratory,
School of Electrical and Information Engineering
University of South Australia,
Adelaide, Australia.

Lakhmi Jain
Knowledge-based Intelligent Engineering Systems Centre,
School of Electrical and Information Engineering
University of South Australia,
Adelaide, Australia.

Colin Fyfe
Applied Computational Intelligence Research Unit,
The University of Paisley,
Paisley, Scotland.

In this paper we discuss the problems associated with a small remnant of seemingly non eradicable accidents in contemporary aircraft landings and propose three autonomous agents whose task it is to jointly monitor the aircraft and its flight crew. Two of these agents will monitor the path of the aircraft, one armed with prior knowledge of how planes tend to land at a particular airport, the other with the ability to extrapolate forward from the plane's current position in order to identify potential dangers. The third agent will monitor the flight crew's behavior for potentially dangerous actions or inactions. These three agents will act together to improve safety in the specific process of landing the aircraft. This paper focuses on the development of the third agent.

Introduction

No one would disagree that air travel, especially in large airliners, has become extremely safe. This has been particularly apparent in the last few decades. The improvement in safety has largely been due to technological innovation in the design of on-board systems and the widespread development and implementation of flight crew training in team management and group effectiveness. Crew resource management (CRM) training has been used by airlines all aver the world in an effort to increase the safety of airline operations. Thatcher (1997, 2000) has suggested that CRM effectiveness could be increased if CRM techniques were introduced much earlier in a pilot's training, at the ab-initio level. This strategy is currently being implemented at the University of South Australia's Aviation Academy. However, given all these advances in aviation safety there remains a statistically constant, and somewhat stubborn, remnant of air crashes which are seemingly not eradicable. Of these accidents, worldwide, Helmreich and Foushee, (1993) have suggested that 70% are due to flight crew actions or in some case inactions. This is despite the fact that pilots are extremely technically competent and well trained in CRM. Pilots undergo regular line checks in both the human factors and technical areas of line operation. Within airlines flight crews are highly trained to operate in the modern cockpit environment. There is consensus that CRM has increased aviation safety. This raises the question as to why these accidents happen and, perhaps even more disturbingly, why they continue to happen, albeit at a very low level of incidence.

In this paper we discuss the problem of continuing accidents in contemporary aircraft approach-and-landings and propose three intelligent software agents whose task it is to jointly monitor the aircraft and its flight crew (Thatcher et al 2004a, 2004b).

The trio of intelligent agents within the proposed paradigm will be organized as follows: Two agents will be physically situated onboard the aircraft. The remaining agent will be physically situated at the destination aerodrome. The agent positioned at the airport will monitor the flight path of the aircraft as it commences its approach. This agent (the Anomaly Detection Agent) will have knowledge of typical aircraft approach profiles for that particular aerodrome.

The other two agents within the paradigm will be situated onboard the aircraft. One of these agents (the Prediction Agent) will have the ability to predict the airplane's future position using its current three dimensional position and vertical and horizontal velocity variables. The predicted future position of the aircraft will be used to identify potential terrain threats. The other agent (the Pattern Matching Agent) also situated onboard the aircraft will monitor the flight crew's behavior and determine if the flight crew are losing situation awareness on the landing approach. This research is in its early stages but the communication and interaction between these three agents is considered essential to the research and future research will concentrate on this area.

This paper will outline the proposed knowledge-based intelligent landing support paradigm with particular emphasis on the third agent, the Pattern Matching agent.

Controlled Flight into Terrain and Approach and Landing Accidents

A controlled flight into terrain accident or CFIT accident can be defined as an accident involving impact with the ground or water by an airworthy aircraft where the flight crew was unaware of the proximity of the ground or water. The majority of CFIT accidents occur on approach and landing and can also be classified as approach and landing accidents (ALA's). However, some CFIT accidents occur in the take off, climb and cruise segments of flight.

Why do such accidents happen and, perhaps more disturbingly, why do they continue to happen?

A Flight Safety Foundation (FSF) report concluded that from 1979 through 1991 CFIT and approach-and-landing accidents (ALAs) accounted for 80% of the fatalities in commercial transport-aircraft accidents (Flight Safety Foundation, 2001) The FSF Approach-and-landing Accident Reduction Task Force Report (Khatwa & Helmreich, 1999) concluded that the two primary causal factors for such accidents were "omission of action/inappropriate action" and "loss of positional awareness in the air".

It seems that most of the CFIT accidents are due to a momentary loss of concentration or awareness during which the flight crew did not consciously notice that a necessary event did not occur, or that an adverse event did occur. Subsequent events are perceived by the flight crew in terms of their current mental model, or awareness, of the situation. Thus it is acknowledged that an event can only be perceived within the framework of the existing paradigm. This is termed situated cognition (Lintern, 1995). Data will continue to be perceived and restructured in terms of the existing mental model until an event happens which forces an unsettling recognition that the pilot's mental model of the world (*weltanschauung*) is actually false. If this happens too late on in a critical process, the result can be an adverse event. This is termed loss of situation awareness.

Situation awareness is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future." (Endsley (1987; 1988)).

One solution to the problem is to increase the level of automation onboard the aircraft. However, automation has associated human factors problems and may lead to a decrease in situation awareness amongst the flight crew. In terms of situation awareness and automation on the flight deck Endsley and Strauch (1997) maintain that "despite their high reliability, accurate flight path control, and flexible display of critical aircraft related information, automated flight management systems can actually decrease" a flight crew's "awareness of parameters critical to flight path control through out-of-the-loop performance decrements, over-reliance on automation, and poor human monitoring capabilities." Further, pilots can in some respects configure the Flight Management System to present a view of the physical world which supports their interpretation of the world or their mental model of the current operating environment. Weiner (1988) describes reports of pilots creating flight paths to wrong locations which went undetected and resulted in collision with a mountain.

We will investigate the critical period associated with ALA's and CFIT accidents when the primary causal factors occur.

Existing Terrain Awareness Technologies

In 1974 the Federal Aviation Administration (FAA) mandated that all heavy airliners be fitted with a Ground Proximity Warning System (GPWS) or Terrain Awareness and Warning System (TAWS). The early GPWS used information from the aircraft's radar altimeter and air data computer to determine the aircraft's vertical distance from the terrain below. The system was somewhat limited because it only perceived in real time the vertical separation between the aircraft and the ground directly below. The Flight Safety Foundation (FSF) CFIT Task Force recommended that early model GPWS be replaced by Enhanced GPWS (EGPWS) or Terrain Awareness and Warning Systems (TAWS) which have a predictive terrain hazard warning function. (Khatwa & Helmreich ,1999).

In response to the report the FAA mandated in 2001 that all heavy transport aircraft be fitted with EGPWS and further, that all turbine aircraft with 10 or more passenger seats be fitted with EGPWS from 2003.

For example, the circumstances which lead to the loss of situation awareness by the flight crew of American Airlines (AA) Boeing 757 that struck a mountain while on descent for a landing at Cali, Colombia, on December 20, 1995 have been investigated (Endsley and Strauch, 1997). Even though the GPWS onboard the aircraft functioned correctly, somewhat surprisingly, it did not help the pilots avoid the collision with high terrain. The reason for this was the operational design parameters of the GPWS

Enhanced Ground Proximity Warning System

The EGPWS compares the aircraft's position and altitude derived from the Flight Management and Air Data computers with a 20MB terrain database. In the terrain database the majority of the Earth's surface is reduced to a grid of 9x9 km squares. Each square is given a height index. In the vicinity of airports the grid resolution is increased to squares of 400m x 400m. The height index and the aircraft's predicted 3 dimensional position 20 to 60 seconds into the future are compared to see if any conflict exists. If it does the EGPWS displays an alert or warning to the flight crew. Other than to initially alert the pilots of "TERRAIN" up to 40-60 s before impact or warn the pilots to "PULL UP" up to 20-30 s before impact it does not offer any other solution to the potential problem.

This research aims to extend the EGPWS by using three intelligent software agents which can plot a course around, or over, possible conflicting terrain and present a solution to the pilot on the cockpit display system or as input to the autopilot.

Intelligent Agents

Wooldridge (2002) describes an intelligent software agent as a program that performs a specific task on behalf of a user, independently or with little guidance. It performs tasks, tailored to a user's needs with/without humans or other agents telling it what to do. To accomplish these tasks, it should possess the characteristics such as learning, cooperation, reasoning and intelligence. By analogy, a software agent mimics the role of an intelligent, dedicated and competent personal assistant. In this application we propose developing three agents, one ground based and the other two aircraft based, which will aid pilots during the critical approach and landing phase. In effect the two onboard agents will act as another flight crew member.

The Anomaly Detection Agent

The anomaly detection agent will be situated on the ground in the air traffic controller centre. Each airport will have its own anomaly detection agent and each agent will be monitored by the local ATC.

A typical airport has many safe landings each day. These are recorded by the air traffic control authorities but not used for automatic sensing of dangerous landings: this is the task of the air traffic controller who has ultimate authority in advising the pilots of potential conflict with terrain or other aircraft. Similar to a human which operates within the cognitive, affective and behavioral domains a software agent can be said to have beliefs, desires and intentions (BDI model). We propose creating an agent using the BDI model (Thatcher et al, 2004a) whose:

- Beliefs are in two major areas: firstly it retains a knowledge of all previously successful landings at that airport. This database itself can be hand-crafted by the (human) air traffic controllers since there may have been some successful landings in the past which, despite being successful, followed a pattern of activity which the air traffic controllers deem to be not good practice. Secondly the agent will have beliefs centered on the current landing – the aircraft's height, horizontal distance from landing strip, speed, heading, lateral distance from landing strip, type of aircraft, weather conditions and any other factors which affect landing performance.
- Desires are that the aircraft lands safely.
- Intentions are to do nothing unless the plane is deemed to be deviating from the historical norm. If such a deviation is noted, the agent informs the air traffic controller who has responsibility for the plane and the pilot himself.

This agent will use anomaly detection as its basic method. Consideration was given to a neural network anomaly detector (e.g. Kohonen's anomaly detector (Kohonen, 1988)) but because it is critical that the warning be given clearly identifying why the warning has been raised, an expert system approach will be used for this application. Thus a series of "if … then … " rules will be created from the database of past successful landings and the current flight's data compared with the rules associated with this database.

The Prediction Agent

Two agents will be situated onboard the aircraft: the Prediction agent will be monitoring the aircraft's position, heading etc and the Pattern Matching Agent (next section) will monitor the pilot's behavior. The Prediction agent is essentially an improved version of the existing EGPWS software described above. The improvements are intended to give a more "intelligent" solution and earlier warning of potential problems than the existing software.

The Prediction Agent has Beliefs about;

- the aircraft's position, heading, speed, rate of descent etc.,
- the landing strip's position,
- weather conditions,
- surrounding ground topology, particularly where dangers are to be found,
- the pilot. This may be controversial to the Pilots' Unions but one must concede that different pilots will tackle tasks differently.

Similarly to the last agent, this agent desires that the plane be landed safely. It again has the intention of doing nothing unless the patterns it is monitoring match potentially dangerous conditions. It might be thought that the Prediction Agent is duplicating the work done by the last agent (Anomaly Detection Agent) but this agent will monitor the descent in a very different manner. The Anomaly Detection Agent will use a database of previous landings *to that particular airport* to ensure that the current landing is *bona fide* and within the parameters of a safe approach. The Prediction Agent will take its knowledge of current position, speed, etc. and knowledge of the local geography to extrapolate the plane's position 5 minutes ahead in order to predict dangerous conditions before they actually occur. This prediction will be derived using an artificial neural network trained with the standard radial basis function methods (Haykin, 1998). A full description of radial basis function networks is given in (Haykin, 1998). The Prediction Agent will be designed as follows: the terrain database derived from the onboard database or from topographical information of the area will be used to generate a grid comprising i=n squares. Each square i at any time t will be assigned variables such as temperature (T(i,t)), atmospheric pressure (P(i,t)), wind component (W(i,t)), terrain height (TH(i,t)), terrain gradient (ΔTH(i,t)), cloud height (CH(i,t)), visability (V(i,t)), and the current aircraft performance variables speed (v(i,t)), altitude (a(i,t)), track (tr(i,t)), and rate of climb (dA/dt(i,t)). A neural network will be trained with this test data until an optimal solution is reached based on the performance of the system using the output measures such as accuracy, sensitivity, false positive and false negative ratios.

If the prediction suggests danger, it is intended that the Prediction Agent will contact the Anomaly Detection Agent and the Pattern Matching Agent. The Anomaly Detection Agent can assert that the current landing pattern is within the recognized safe zone but if it deemed to be close to the edges of this zone, an alert will be issued to the pilot and the air traffic controller.

The alert to the Pattern Matching Agent will be discussed in the next section.

The Pattern Matching Agent

The Pattern Matching Agent will also be based on the BDI model and has beliefs about:

- The recent past behavior of the pilot
- Typical behaviors of the current pilot
- Behaviors which are typical of pilots losing situational awareness, performing an inappropriate action or not performing an appropriate action.

Again its desires are that the plane lands safely and its intentions are to do nothing unless it matches the pilot's current behavior with dangerous practice.

The Pattern Matching Agent will be equipped with a database of behaviors which are suggestive of, or a prelude to, the loss of situation awareness. In other words, this agent will fulfill the role of a dedicated professional who, sitting in the cockpit, would identify the pilot's actions (or inactions) as worthy of concern. This pattern matching is accomplished by a simple Associative Artificial Neural Network (Haykin, 1998) which matches approximately existing patterns of behavior to those in the database.

There is a body of research that indicates that pattern matching together with schema and mental models facilitates the development of situation awareness (Kaempf et al, 1993). Endsley and Bolstad (1994) found evidence that fighter pilots with higher levels of SA had better pattern matching skills. Kaempf et al (1996) discovered that pattern matching to situation prototypes accounted for 87% of decisions by tactical commanders.

This research aims to develop a database of typical pilot behaviors or actions during an approach. Further data will be included in the database to describe rate of descent, speed, flap setting, speed brake armed, altitude etc. Perhaps this database could be described as a mental model of a descent? It might be difficult to envisage a software agent having a mental model of a typical approach. But we should consider that a software agent (Pattern Matching

Agent) could have a database of typical pilot behaviors that could be compared with the actual behaviors. How big the database has to be, for the agent to be considered as processing a mental model is a question for future research. However, for this agent to function in the cockpit and communicate potential hazardous behaviors, be they actions or inactions, the agent must develop SA within its beliefs, desires and intentions model.

All three agents will have the ability to communicate with each other at all times. To this extent each agent will have beliefs about the other two. When the Pattern Matching Agent received a warning from either of the others, it would respond with a degree of confidence about the pilot's current situation awareness. We currently intend the Pattern Matching Agent as a reinforcement mechanism for the other two agents. At this stage in the research we do not envisage this agent overruling warnings communicated by the other two. Further, the combination of the three agents would achieve the three levels of SA (Endsley, 2000)) Level 1 – Perception, Level 2 - Comprehension and Level 3- Projection.

Conclusion

We have identified the specific process of approach-and-landing accidents as one which might successfully be augmented with intelligent agent technology. We thus have proposed three agents:

1. The first will be situated on the ground and will have a knowledge of typical landings at the current airport.
2. The second will be situated onboard the aircraft and will be attempting to use the aircraft's current position and heading and knowledge of the local geography to predict potential dangers.
3. The third will be also onboard the aircraft and will be monitoring the behavior of the flight crew for actions indicative of the loss of situation awareness.

This research is in its early stages but we consider the interaction between these three agents to be central to the research and future research will concentrate on communication between the three agents.

References

Endsley, M.R., (1988) Design and evaluation for situational awareness enhancement. *Proceedings of the 32nd Annual Meeting of the Human Factors Society*, 97-101.

Endsley, M. R. (1988). Situation Awareness Global Assessment Technique (SAGAT). *Proceedings of the National Aerospace and Electronics Conference (pp. 789-795).* New York: IEEE.

Endsley, M. R., (2000), Theoretical Underpinnings of Situation Awareness: A Critical Review. In Endsley, M. R. and Garland D. J., (Eds.). (2000)., *Situation Awareness Analysis and Measurement*. Mahwah, NJ: Lawrence Erblaum Associates.

Endsley M. R.. and Bolstad, C. A. (1994). Individual diffences in pilot situation awareness. *International Journal of Aviation Psychology*, 4(3), 241-264.

Endsley, M. & Strauch, B. (1997) Automation and Situation Awareness: The Accident at Cali, Columbia, In the *Ninth International Symposium on Aviation Psychology*, Columbus, OH

Haykin, S. (1999) *Neural Networks: a Comprehensive Foundation.* Prentice Hall International.

Helmreich, R. L. & Foushee, H. C (1993) Why Crew Resource Management? In Weiner, E. L., Kanki, B. G. & Helmreich, R. L., (Eds.), (1993). *Cockpit Resource Management*. San Diego: Academic Press.

Kaempf, G. L., Klein, G. A., Thordsen, M. L. and Wolf, S. (1996). Decision making in complex navel command and control environment. *Human Factors*, 38(2), 220-231.

Kaempf, G. L., Wolf, S. and Miller, T. E. (1993). Decision making in the AEGIS combat information center. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (pp.1107-1111). Santa Monica, CA: Human Factors and Ergonomics Society.

Khatwa & Helmreich ,(1999), Analysis of Critical Factors During Approach and Landing in Accidents and Normal Flight., Data Acquisition and Analysis Woking Group, Flight Safety Foundation Approach-and-landing Accident Reduction Task Force. *Flight Safety Foundation, Flight Safety Digest*, Nov 1998-Feb 1999.

Kohonen, T. (1988). *Self-organization and associative memory*, Springer-Verlag., Berlin

Lintern, G (1995). Flight Instruction: The Challenge From Situated Cognition. *The International Journal of Aviation Psychology*, 5(4), 327-350

Thatcher, S. J. (1997). Flight Instruction or Flight Facilitation: A Foundation for Crew Resource Management. In the *Proceedings of the Ninth International Symposium on Aviation Psychology*, Columbus, OH.

Thatcher, S. J. (2000). The foundations of crew resource management should be laid during ab-initio flight training. In Lowe, A. R. &

Hayward, B. J., (Eds.) (2000), *Aviation Resource Management*. _Aldershot, England: Ashgate

Thatcher, S. J., Jain, L. C. and Fyfe, C., (2004a). "An Intelligent Aircraft Landing Support System." In *Lecture Notes in Computer Science: The Proceedings of the 8th International Conference on Knowledge-based Intelligent Information & Engineering.*, Berlin: Springer-Verlag, 2004

Thatcher S J., Jain, L. C. and Fyfe, C., (2004b). Theoretical Framework for a Knowledge-based Intelligent Enhancement to an Aircraft Ground Proximity Warning System. In Negnevitsky, M. (ed.), *The Proceedings of the 2nd International Conference on Artificial Intelligence in Science and Technology (AISAT'2004.)* (2004), Tasmania:University of Tasmania

Wiener, E. L. (1988). Cockpit automation. In E. L. Wiener & D. C. Nagel (Eds.), *Human Factors in Aviation* (pp. 433-461). San Diego: Academic Press.

Wooldridge, M., *An Introduction to MultiAgent Systems*, Chichester: John Wiley & Sons Ltd, 2002.

EFFECTS OF CDTI DISPLAY DIMENSIONALITY AND CONFLICT GEOMETRY ON CONFLICT RESOLUTION PERFORMANCE

Lisa C Thomas
Christopher D Wickens

Institute of Aviation
University of Illinois
Savoy, IL

With the presence of a CDTI that provides graphical airspace information, pilots can use a variety of conflict resolution maneuvers in response to how they perceive the configuration of the conflict. However, across previous studies on conflict resolution using CDTIs, there has been little apparent consistency in maneuver safety, flight axis preferences (lateral or vertical), or turning direction within a flight axis. These inconsistencies may be due to a limited range of conflict geometries and/or display frames of reference. This article describes a study that incorporates three displays with different frames of reference and a wide range of conflict geometries to determine their specific effects on maneuver preferences. Results indicated that the designs of the two 3-D displays, which included features to reduce spatial ambiguities, produced performance levels nearly equivalent to the 2-D coplanar display in almost all conflict geometry conditions. Overall, display dimensionality had no effect on success or response times and only a limited effect on direction preference within the lateral axis. Conflict geometry, especially lateral approach angle, affected success, response times, and preferences for maneuvering along different flight axes.

Introduction

The move towards Free Flight will require that the pilot have access to displays that will accurately support his/her understanding of the airspace (RTCA, 1995). This understanding will contribute to the pilot's ability to navigate through the airspace, maintain self-separation, and resolve potentially risky flight situations (conflicts) as they arise. In support of these proposed new responsibilities, cockpit displays of traffic information (CDTIs) are being developed to support pilot awareness and understanding of the airspace and the traffic within it (e.g. Johnson, Battiste, & Bochow, 1999). Currently, FAA-certified CDTIs do not contain any features for directly interacting with the flight plan to (for example) resolve conflicts, but there has been research to investigate the effectiveness of such a feature.

A conflict is defined as the loss of the minimum required separation distance between two aircraft, defined in this study as 5 nautical miles lateral and ± 1000 ft vertical; a resolution involves creating a new flight path that ensures that the two aircraft do not lose minimum separation. Although the Federal Aviation Regulations (FAR 91.113) recommend that conflicts be resolved by making lateral changes only, it is not mandatory and in fact the on-board Traffic Collision Avoidance System only provides vertical resolution recommendations when a conflict is detected. With the presence of a cockpit-based display that provides more detailed 3-D airspace information, pilots have more opportunities to use a wide variety of conflict resolution maneuvers (airspeed, altitude, and/or heading changes) in response to how they perceive the configuration of the conflict. Thus, it is also important to establish the extent to which the CDTI induces resolution maneuvers that are consistent or inconsistent with either TCAS resolution advisories (vertical only), or FAA "rules of the road" (lateral only).

Display Dimensionality

A CDTI is designed to show air traffic from the perspective of the pilot's own aircraft ("ownship;" Johnson, Battiste, & Bochow, 1999). Understanding of the airspace is supported to varying degrees by the dimensional frame of reference of the CDTI. The frame of reference dictates how the spatial information is depicted, whether it is two- or three-dimensional. It has been well established that for each frame of reference, there are benefits as well as costs (see Wickens, 2002, 2003 for review). Thus, selecting the most appropriate CDTI frame of reference depends on identifying the benefits and costs of each type of display for the particular tasks facing the pilot (e.g. conflict detection, resolution).

For example, in a 2-D coplanar display, there are two orthographic views of the airspace (from above and from the side or behind ownship) showing only two dimensions each. Each view in this format shows absolute spatial information in two dimensions without ambiguity, but requires effortful cognitive integration across both views for a full understanding of the 3-D environment (Wickens, 2002, 2003).

In a 3-D perspective display, all three dimensions are integrated and displayed in a manner analogous to the environment being depicted, but the particular viewpoint will cause some distortion in at least one, and possibly all three, spatial dimensions. Line-of-sight ambiguity is a result of the projection of the 3-D environment onto a 2-D screen, and can produce biases in estimating distances, such as foreshortening, along the compressed axes (Wickens, 2002).

One potential solution to the problem of viewpoint-related foreshortening is to allow the viewpoint to be positioned in a variety of angles so that the 3-D spatial environment may be viewed from different directions that disambiguate the relevant spatial information. This can be accomplished by either providing several pre-set viewpoints, which the pilot may choose between, or by allowing the pilot to freely and continuously reposition the viewpoint as desired (e.g. Wickens & Helleberg, 1999). Determining the feasibility of resolving this ambiguity through viewpoint rotation is one of the objectives of the current study.

Conflict Geometry

The conflict geometry between two planes (ownship and "intruder") can be defined by three parameters: **altitude** of intruder (both absolute and relative to ownship), **airspeed** of intruder (both absolute and relative to ownship), and the **angle** formed by the intersection of the trajectories of the two aircraft. Conflict geometry has been found to affect the type of conflict avoidance maneuver chosen by the pilot and the safety of those maneuvers (e.g. Scallen, Smith, & Hancock, 1997; Johnson, Bilimoria, Thomas, Lee, & Battiste, 2003).

Maneuver Choice Summary

The collective findings of the influences of conflict geometry and display dimensionality on maneuver choice and maneuver safety are somewhat inconsistent, but do allow a few conclusions to emerge, with varying degrees of certainty, as summarized in papers by Wickens, Helleberg, & Xu (2002), Alexander, Wickens, & Merwin (2005), and Thomas & Wickens (2005, in preparation).

Regarding maneuver choice, there is a general tendency to choose vertical over lateral maneuvers, at least when light (e.g., GA) aircraft simulations are involved. Furthermore there appears to be a tendency for the vertical preference to be enhanced to the extent that the linear vertical representation of the co-planar display is present. That is, the coplanar display enhances the vertical preference, while the 3D display diminishes it.

Within the vertical dimension, there appears to be a climbing preference that emerges with, and is consistently shown by, the co-planar display (Wickens & Helleberg, 1999; Alexander et al, 2005, Experiments 1 and 2; O'Brien & Wickens, 1997). However this preference is reduced, and sometimes reversed, with a 3D display which sometimes invites more descents than climbs (Alexander et al, 2005, Experiments 1 and 3).

The pattern of climb-descent preference is somewhat complicated by the influence of conflict geometry. With the coplanar display, pilots generally chose to maneuver in the opposite direction of the vertical behavior of non-level traffic. That is, they climb when it descends and vice-versa. However, with the 3D display this "vertical opposite tendency" appears to be less consistently manifest, and is sometimes replaced by a tendency to maneuver vertically in the same direction as the traffic (O'Brien & Wickens, 1997; Alexander et al, 2005, Experiment 1). Finally, at least within the coplanar display, the overall climbing tendency appears to be amplified to the extent that traffic approaches from the front (head-on conflicts).

Maneuver Safety Summary

Regarding maneuver safety, as typically measured by the amount of time during which there is a predicted loss of separation, whenever safety differed between display formats, this measure favored the coplanar display. Such a difference may be attributable to the ambiguity of the 3D display because such 3D costs tended only to emerge when the traffic was non-level (climbing or descending), a circumstance that will leave its vertical trajectory ambiguous on the 3D but not the 2D display (Alexander et al, 2005, Experiment 1; O'Brien & Wickens, 1997). This effect is replicated on Air Traffic Control displays as well (Wickens, Miller, & Tham, 1996).

Our study was specifically designed to contrast a 2D co-planar display with two versions of a multi-viewpoint 3D CDTI that were both designed to address the 3D ambiguity problems that have plagued single viewpoint 3D displays in the past (Alexander et al, 2005; Wickens & Helleberg, 1999). Both 3D CDTI versions allow for pilot control over the viewpoint and provide continuous motion between viewpoint rotations, producing motion parallax and different perspectives which may reduce the spatial ambiguities of any one perspective.

Hypotheses

H1. 3D ambiguity will be manifest as a drop in the success rate for resolving conflicts using the two 3-D CDTIs compared to the 2-D coplanar.
H2. However, the fact that alternative viewpoints are provided for each of the two 3-D CDTIs may be sufficient to reduce, and perhaps eliminate, the 3D costs relative to previous experiments.
H3. There will be a general preference for vertical maneuvering over lateral.
H4. Further, the better (more precise) rendering of the vertical axis in the 2-D co-planar CDTI may amplify this preference. That is, the 2-D co-planar CDTI will cause more vertical maneuvering than either 3D CDTI.
H5. Within the vertical axis, climbs will be preferred over descents, and vertical maneuvering will be opposite the traffic maneuvering, at least [particularly] with the coplanar displays.
H6. Vertical geometry (climbing-descending traffic) will present more difficult conflicts to resolve because this involves 3 simultaneous axes of change, and may lead to less accurate resolutions and/or longer response times in creating resolutions.

Methods

Participants. Thirty student pilots from the University of Illinois participated in this study and were reimbursed for their participation.

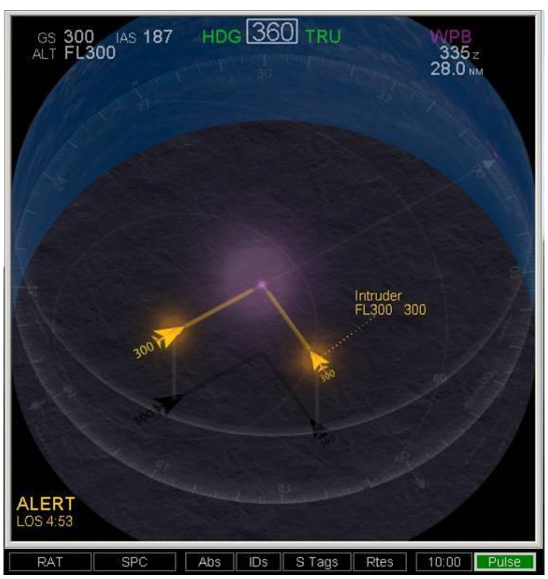


Figure 1. *Cockpit display of traffic information. Image shows Toggle condition View 1, conflict with an intruder in a crossing conflict from the left and traveling at the same speed and same altitude as Ownship.*

Displays. There were three different CDTI formats. The Coplanar display consisted of two side-by side views of the airspace, one top-down (showing the horizontal plane) and one from the side (showing vertical information). The Toggle display consisted of one display with two 3D viewpoint options that could be switched by pressing the “View 1/View 2” button on the CDTI button bar (see Figure 1). View 1 was set to 60 degrees elevation and 330 degrees azimuth (above and slightly to the left of ownship), and View 2 was set to 30 degrees elevation and 60 degrees azimuth (slightly above and far to the right of ownship). The Manipulable display consisted of one display with a viewpoint that could be set anywhere in the vertical range of 0°-90° or laterally from 0°-360° when moved around by the participant using the mouse.

Design. Table 1 outlines 54 (or 3 x 3 x 3 x 2) unique conflict geometries (the within subject variables) which were used as the conflict trials. The total number of experimental trials was tripled by choosing three angles within each subset of conflict angles so that the three sets of conflicts were similar but not identical. The conflict geometries were defined by the intruder’s position relative to ownship and covered the spectrum of conflicts that may occur in real flight. The same set of 162 trials were presented in a randomized orders to different pilots in each of the three display conditions. All conflicts were direct collision courses between ownship and intruder.

Between	**Within**		
Display	**Angle**	**Altitude Change**	**Relative Speed**
Coplanar	Head-on (150º-210º)	Ascending to OS’s alt.	Faster than ownship
Toggle	Crossing (70º-110º, 250º-290º)	Level, same altitude as ownship	Same as ownship
Manipu-lable	Overtake (20º-50º, 310º-340º)	Descending to OS’s alt.	Slower than ownship

Table 1. Each level of each variable in this study.

Procedure and Tasks. At the beginning of the experiment, the pilots were quasi-randomly placed into one of the three display conditions. After signing consent forms, reading instructions, and performing practice trials, each pilot viewed 162 experimental trials consisting of a conflict between ownship and an intruder. Each trial was constructed so that the conflict was predicted to occur in 5 minutes from the start of the trial. Hence, head-on conflicts started with larger separation, and closed at a faster speed. Pilots were instructed to resolve each of the conflicts by making one of the following resolution maneuvers: lateral (by using the mouse to click and drag ownship's flight path into a new configuration), vertical (by using the mouse to click up/down arrows in a pop-up altitude change menu), or both in combination. Feedback from the alerting color changes indicated whether a proposed resolution was successful: if it was, the color of the aircraft changed away from yellow. Once a successful resolution was entered, pilots clicked an Enter and an Execute button, and after 5 seconds the next trial began.

Results

Resolution Success

Success data were first calculated by the percentage of the 18 trials within each category of display condition x conflict angle x relative speed (collapsed across altitude, after it was determined that altitude had no main effect or interactions with other variables) where participants entered a successful resolution. The data were skewed, and therefore transformed using the arcsine transformation A repeated measures ANOVA was conducted on the data, using the three levels of conflict angle and three levels of relative speed as the repeated measures, and display condition as the between subjects variable.

There was no significant main effect of display type on success of resolution. The significant main effect of conflict angle on resolution success ($F_{2, 54}$ =5.70, p<0.006) shows a decrease in performance as the conflict angle gets smaller: conflicts with intruders approaching at head-on conflict angles are easier to resolve (96%) than crossing angles (93%) or overtake angles (93%).

The significant interaction between display condition and conflict angle shown in figure 2 ($F_{4, 54}$ = 3.01, p<0.026) reflects the fact that only in the Manipulable condition were crossing and overtaking conflicts more difficult to resolve.

A significant main effect of relative speed of the Intruder compared to ownship ($F_{2, 54}$ = 6.18, p<0.004) reflects the fact that the best resolution performance occurs when the Intruder aircraft is faster (96%) than Ownship (no difference between same [93%] and slower [94%] speed performance).

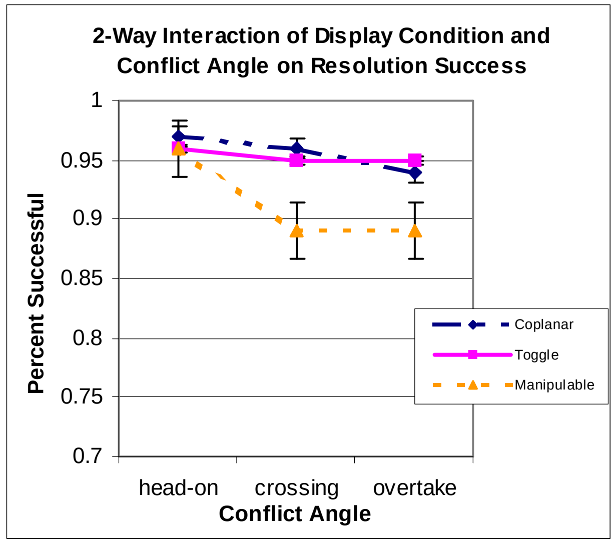


Figure 2. *Resolution success as a function of display condition and conflict angle.*

Response Times

There was no significant main effect of display on response times nor was there any interaction between display and any parameter of conflict geometry.

There was a significant main effect of conflict angle on response times ($F_{1.6, 44.1}$ =36.32, p<0.001. Head-on (16.3 s) conflicts were the fastest to resolve (and also the most successful; refer to Figure 2), then crossing (18.1 s), and overtake (19.8 s) were the slowest (and least successful) to resolve. This is likely due to the fact that in head-on conflicts, the perceived time pressure from the faster closure rate may encourage faster maneuver selection.

Maneuver Axis Preference

Maneuver axis type for each successful resolution was categorized as one of three types: lateral (i.e. turn toward or away), vertical (i.e. climb or descend), or dual-axis combination (e.g. climbing left turn) maneuvers. Frequency of maneuver axis types was determined by calculating proportion of each type of maneuver across successful trials in each conflict geometry category for each pilot. These frequencies were then analyzed to determine whether pilots demonstrated a preference for one maneuver axis type over another.

An initial analysis revealed a slight but statistically non-significant (p=0.14) preference between the three maneuver categories (vertical: 40%, lateral 29%, combined 30%). To determine how other aspects of the conflict geometry might have moderated this preference profile, the three categories of maneuvers were then classified as a second independent variable in an ANOVA, so that the profile modification would be revealed as a statistical interaction between axis choice, and other display/geometry variables.

The analysis revealed that there was a significant interaction between conflict angle and maneuver axis preference ($F_{4, 104} = 13.06$, $p<0.001$; refer to Figure 3, left graphs). Vertical maneuvers were preferred over both lateral and combination in crossing conflicts, and over lateral in overtake angle conflicts. There was a significant interaction between altitude change and maneuver axis preference ($F_{4, 104} = 3.36$, $p<0.012$; Fig 3, center graphs). Vertical maneuvers were preferred over lateral and combination when the intruder was flying level. There was a significant interaction between relative speed and maneuver axis preference ($F_{4, 104} = 8.49$, $p<0.001$; Fig 3, right graphs). Vertical maneuvers were preferred over lateral and combination when the intruder was flying faster or at the same speed as ownship.

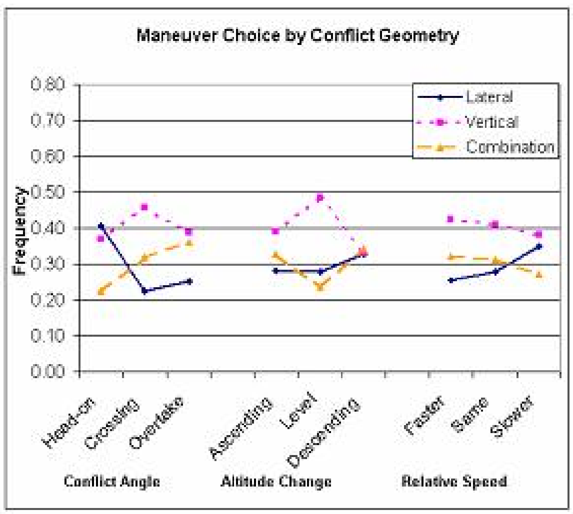


Figure 3. *Main effects of each parameter of conflict geometry on maneuver flight axis preference.*

Within-Axis Direction Preference

Maneuver axis preferences were further analyzed to determine whether there were preferences for turning one direction or the other within an axis.

Single-axis vertical direction preferences (ascents vs. descents) were evaluated on the basis of the intruder's vertical behavior (ascending or descending only; cases where intruder was flying level were analyzed separately). The vertical choice was significant (F1, 26 = 50.4, $p<0.001$), and shows a strong preference to maneuver vertically away from the intruder across all display conditions. This is consistent with the pattern of prior research using the 2-D coplanar CDTI research, but is less consistent with the pattern using the 3D display.

Vertical direction preferences when the intruder was flying level were then evaluated. There was a marginally significant effect of vertical direction preference ($F_{2, 52} = 2.84$, $p<0.07$), where ascents were most preferred, followed by level flight and then by descents, consistent with the results in Wickens et al (2002). Neither the display condition nor any dimension of conflict geometry had a significant effect on vertical direction preference.

Discussion

There appeared to be some support for H1, in that the Manipulable display produced the poorest performance overall and specifically in cases where the conflict was in a crossing over overtake configuration. However, there was also support for H2: there was no performance difference between the 2-D Coplanar and 3-D Toggle conditions, and even the 3D costs of the Manipulable display were attenuated, only manifest in two of the three conflict angle conditions, and then only in a 6% loss of accuracy (Figure 2).

There was marginal support for H3: overall, there was a slight preferences for vertical maneuvers, and in particular, vertical maneuvers were preferred over lateral and combination in both crossing and overtake angle conflicts, when the intruder was flying level, or when the intruder was flying faster or at the same speed as ownship. There were no circumstances in which lateral maneuvers were preferred.

Display dimensionality had no significant effect on maneuver axis preference; contrary to H4, the three-dimensional nature of the displays did not appear to significantly alter maneuver flight axis preferences compared to the 2-D coplanar display, and again, as with Hypothesis 2, suggesting that the interactivity of the viewpoints attenuated the previously-observed influences of the 3D display..

There was limited support for H5 in the data indicating that ascents were preferred over descents when the intruder was flying level. Furthermore, the preference to maneuver vertically away from the intruder (choosing descents significantly more often than ascents when the Intruder was descending) was stronger than the general preference for ascending maneuvers over descending. There was no support for H6: neither success rate nor response time was affected by whether the Intruder was flying level or making a vertical change in either direction.

One possible reason why our results did not replicate some previous findings of differences between 2-D and 3-D CDTIs is that the multi-viewpoint designs of the 3-D CDTIs used in this study reduced the ambiguities associated with single-viewpoint 3-D CDTIs, and thus reduced the decrements in performance attributed to those ambiguities. In addition, in our experimental paradigm pilots were asked to plan a resolution maneuver with interactive tools that provided safety feedback, but were not required to carry it out. Previous studies had pilots actually fly their resolution maneuvers, and it is unclear what differences, if any, may exist for maneuver flight axis preferences between planning a theoretical resolution and using a flight simulator to carry one out.

Conclusions

The interactive features of the two 3-D displays (multiple set viewpoints, continuously manipulable viewpoint) appeared to reduce ambiguity and produced success and response time performance levels more or less equivalent to the non-interactive 2-D coplanar condition, with a limited time cost associated with the Toggle feature. So far, analysis of the results indicates that there are limited main effects of display dimensionality and conflict geometry on conflict resolution characteristics.

Acknowledgments

This research was sponsored by NASA Ames Research Center #NAG 2-1535. Dr. David Foyle was the technical monitor. The authors also thank Dr. Walt Johnson and Dr. Sandy Hart for their contributions, and Dominic Wong, George Lawton, and Sharon Yeakel for their software development expertise.

References

Alexander, A. L., Wickens, C. D., & Merwin, D. H. (2005). Perspective and coplanar cockpit displays of traffic information: Implications for maneuver choice, flight safety, and mental workload. *Int'l Journal of Aviation Psychology, (15)*, 1-21.

Ellis, SR, McGreevy, MW, & Hitchcock, RJ. (1987) Perspective traffic display format and airline pilot traffic avoidance. *Human Factors*, 29, 371-382.

Johnson, WW, Battiste, V, & Bochow, S. (1999) A cockpit display designed to enable limited flight deck separation responsibility. *1999 WAC Proceedings*. Anaheim, CA.

Johnson, WW, Bilimoria, KD, Thomas, LC, Lee, HQ, & Battiste, V. (2003) Comparison of human and automation generated conflict resolutions. *AIAA GNC Conference, August 2003.*

O'Brien, J.V., & Wickens, C.D. (1997). Free flight cockpit displays of traffic and weather: Effects of dimensionality and data base integration. *41st Annual HFES Meeting Proceedings*. Santa Monica, CA: Human Factors & Ergonomics Society.

RTCA (1995) *Report of the RTCA Board of Directors' Select Committee on Free Flight*. Washington, CD: RTCA, Inc.

Scallen, Smith, K., & Hancock, P.A. (1997) Influence of color cockpit displays of traffic information on pilot decision-making in free flight. *41st Annual HFES Meeting Proceedings.* Santa Monica, CA: Human Factors & Ergonomics Society.

Thomas, L.C., & Wickens, C.D. (2005, in preparation) Display dimensionality and conflict geometry effects on conflict detection and resolutions. *ARL Tech Report.*

Wickens, C.D. (2002) Spatial awareness biases. *ARL Tech Report*.

Wickens, C.D. (2003), Aviation Displays. In P. Tsang and M. Vidulich. *Principles and Practices of Aviation Psychology*. Mahwah NJ. Erlbaum

Wickens, C.D., Helleberg, J., and Xu, X. (2002) Pilot maneuver choice and workload in free flight. *Human Factors, 44*, 171-188.

Wickens, C.D., & Helleberg, J. (1999) Interactive perspective displays for airborne hazard awareness. *ARL Tech Report*

Wickens, C.D., Miller, S., & Tham, M. (1996). The implications of datalink for representing pilot request information on 2D and 3D air traffic control displays. *International Journal of Industrial Ergonomics, 18*, 283-293.

DISORIENTATION IN VFR PILOTS: FLIGHT PERFORMANCE AND PSYCHOPHYSIOLOGICAL CHANGES DURING A FLIGHT SIMULATOR TRAINING

Karin Tropper & **K. Wolfgang Kallus**
Department of Psychology, Karl-Franzens-University of Graz
Graz, Austria

Disorientation due to flying into instrument meteorological conditions (IMC) is a major safety hazard for VFR pilots (VFR: visual flight rules) as confirmed by aviation accident databases. The objectives of our research are the development and evaluation of systematic training programs to cope with different kinds of disorientation phenomena and the analysis of the psychophysiological processes during dis- and reorientation. A study was conducted using the multi-axial moveable flight simulator DISO (AMST Systemtechnik GmbH, Austria). 25 pilots were randomly allocated to one of three testing groups (one control- and two experimental training groups). The flight performance data confirm that participants with a training show better performance data in a test phase than pilots without training. The simulation scenarios are of high impact: Heart rates are clearly increased in response to more demanding segments of flight as e.g. during takeoff and landing. Analyses within the test profile "unusual-attitude recovery" demonstrate – in addition to the expected increase of heart rate due to higher mental workload – an important interaction: The increase is lower for pilots having received an unusual-attitude recovery training. First EEG results illustrate changes in the alpha- and beta band due to changing strain. To sum up, this study tries to make a contribution to basic research by analyzing psychophysiological processes as well as to applied science by emphasizing the importance and effectiveness of orientation training programs for VFR pilots.

Introduction

Disorientation due to flying into instrument meteorological conditions (IMC) is a major safety hazard for VFR pilots. Analyses of aviation accident databases confirm that in general aviation fatal aviation accidents are often classified as involving visual flight rules (VFR) into instrument meteorological conditions (e.g. Goh & Wiegmann, 2001; Véronneau & Evans, 2004).

Our concept to explain spatial and geographic orientation and disorientation bases on the model of anticipatory action regulation from Hoffmann (1993) and the model of situation awareness (SA) from Endsley (2000). "Situation Awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1995, p. 65). Situation awareness involves a correct appreciation of many conditions. The most relevant aspects in aviation are three-dimensional spatial awareness, system (mode) awareness, and task awareness (Wickens, 2002). As correct orientation is a central factor of situation awareness, loss of orientation leads to loss of situation awareness (LSA).

The objectives of our research are the development and evaluation of systematic training programs helping to cope with different kinds of disorientation phenomena, using the multi-axial moveable (continuous yaw, limited pitch and roll) flight simulator DISO (Disorientation Trainer, AMST Systemtechnik GmbH, Austria).

In a first study, 26 jet pilots participated. The main results are that the simulator illustrates disorientation phenomena very realistically, that flight performance increases after a disorientation recovery training, and that worse performance in simulator exercises – e.g. crash during the profile "Black hole approach" – is accompanied by high physiological stress as indicated by increases in heart rate (Kallus & Tropper, 2004). Based on these results, a study was designed with adopted profiles for VFR pilots (Haug, 2003) using again the multilevel multi-method approach for the evaluation of the training effects and the analysis of cognitive, psychological and psychophysiological processes.

Method

Design and Subjects

25 VFR pilots (average age of 43 years, SD = 10.5, 23 men, 2 women, all owning a private flight license) were randomly allocated to one out of three testing groups. The experimental design is given in Table 1. Table 2 shows an overview of the flight profiles. Every participant completed three phases in the flight simulator. The nine pilots of the training group attended the awareness training ("awareness") during phase I, followed by the training with orientation- and unusual-attitude recovery exercises (phase II, "training"). The eight pilots of the awareness group also went through the awareness phase, but instead of the training phase they completed the control condition "free flight". The control group (n = 8) went

through two free flight phases instead of the training. All 25 pilots passed the test (phase III) at the end of the testing day. The simulator exercises were based on a PC7 simulation.

Table 1. *Experimental design*

	PHASE I	PHASE II	PHASE III
TRAINING GROUP (n = 9)	Awareness	Training	TEST
AWARENESS GROUP (n = 8)	Awareness	Control condition II	TEST
CONTROL GROUP (n = 8)	Control condition I	Control condition II	TEST

Table 2. *Overview of the simulator profiles*

PHASE I	
AWARENESS	**CONTROL CONDITION I**
Cockpit Instruction	*Cockpit Instruction*
Instruction flight at excellent weather conditions (WX)	*Instruction flight at excellent weather conditions (WX)*
VFR flight at min. WX, mountains	**Free Flight I**
VFR flight, mountains, clouds tilt	
Passive spin profiles: Gyrospin I and Gyrospin II	*Passive spin profiles: Gyrospin I and Gyrospin II*

PHASE II		PHASE III
TRAINING	**CONTROL CONDITION II**	**TEST**
VFR flight at min. WX, mountains, visual and VOR	**Free Flight II**	**VFR flight at min. WX, mountains**
VFR flight at min. WX, mountains, Radar Vectors		**Unusual-attitude recovery**
Unusual-attitude recovery training		

Instruction Flight. The instruction flight takes place under conditions of good visibility (about 80 km). It leads the pilot along a standardized flight path with the takeoff at Kalamata (Greece), leading to the coast, along the coast, briefly across the sea, into terrain with mountains and finally back to the airport of Kalamata. After passing the last of five turning points and before landing in Kalamata, the pilot flies two maneuvers: an aileron roll and a looping. The flight path is approximately 33 nautical miles long and it takes about 18 minutes to fly the whole circuit (including takeoff, the flight maneuvers and landing). As aid, the pilot gets a colored map of Peloponnes into which the flight path is drawn. Additionally, standardized headings are used by the instructor pilot to lead and help the pilot via radio connection. The instructor pilot also took on the tasks of an air traffic controller.

Test profile VFR flight at minimal weather conditions, mountains. This profile begins with conditions of bad visibility (10 km). It is planned to fly the same route as during the instruction flight and the pilot is explicitly instructed “to behave as in a real flight situation”. The visibility deteriorates further with time (5 km). It is overcast and the mountains are in clouds. It is not possible to fly the whole planned circuit under VFR condition. Because visibility deteriorates gradually, it is expected that not all pilots become aware of the hazard and use visual flight rules into instrument meteorological conditions.

Test profile Unusual-attitude recovery. Unusual-attitude recovery means the process of returning the aircraft to near straight and level from an unexpected bank and / or pitch angle. The exercise is drawn from jet pilots´ training courses. At the beginning of this profile, the PC7 is already airborne. After about two minutes, the instructor pilot takes over the control of the PC7 and sets certain – standardized – flight parameters via the external workstation. During the set-up time the participant inside the flight simulator keeps his eyes closed. After taking over the control from the instructor pilot, the pilot in the simulator is required to reach safe flight parameters (to recover) as fast as possible. This exercise is conducted ten times.

Procedure

The examinations lasted five to eight hours per pilot. Before and after each flight simulator phase, a two minute resting measurement (baseline, eyes closed) was conducted. After each phase (outside the simulator), the pilot took part in an extensive reconstruction interview concerning the flight profiles.

Dependent Variables

Aviation performance (observation data, instructor pilot ratings, time-measurements), psychological data (questionnaires for analyzing changes in subjective physical and psychical state, reconstruction interviews), and physiological variables (ECG, EEG, EOG, EDA) were measured.

Some results concerning the following dependent variables are reported here:

- Flight performance: observation data
- ECG: heart rate – deviation from baseline: Positive differences signify an increase in heart rate in comparison to the resting measurements.
- EEG: spontaneous activity

EEG was recorded by eight bipolar channels (positions of electrodes cf. Table 3; the ground electrode was fastened to the forehead). The electrode impedances were below 5 k ohms and the sample rate was 128 Hz. Recorded data were subject to visual inspection using the BrainVision software package of the Company Brain Products GmbH (Munich). Seconds with artefacts were excluded from further analyses. The EEG from 1 second periods were submitted to spectral analysis using the Fast Fourier Transformation (full power spectra, Hanning window). After averaging the absolute power values of the 1 seconds periods of certain sections of measurement, the data were combined to the standard bands of alpha (8-13 Hz) and beta (14-30 Hz).

Table 3: *Positions of the 16 EEG electrodes (eight bipolar channels, frontal to occipital regions*)

Channel 1	F3 - FC´3
Channel 2	F4 - FC´4
Channel 3	FC3 - PC3
Channel 4	FC4 - PC4
Channel 5	C3 - P3
Channel 6	C4 - P4
Channel 7	P´3 - O1
Channel 8	P´4 - O2

The rate of missing EEG-data is beyond five percent for each channel (due to continuously artifacts – e.g. muscle activity – or technical problems). No missing EEG data have been replaced and to lose no additional data, only univariate analyses (power of only one channel) have been calculated.

Results

Flight Performance

During the test phase (VFR flight at minimal weather conditions, mountains), the pilots of the control group caused the highest number of crashes [Pearson-Chi^2 (df=2, n=25) = 10.96, p = .004, Figure 1]. Figure 2 illustrates that pilots of the training group show the tendency to enter the cloud layer less often than participants of the other two groups [Chi^2 (df=2,n=25) = 4.99, p = .102].

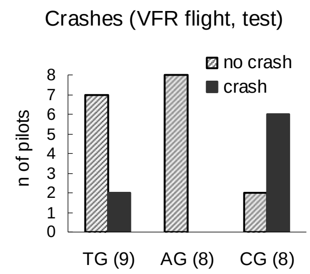


Figure 1. *Crashes during the VFR flight of the test phase separate for pilots of the training group (TG), awareness group (AG), and control group (CG).*

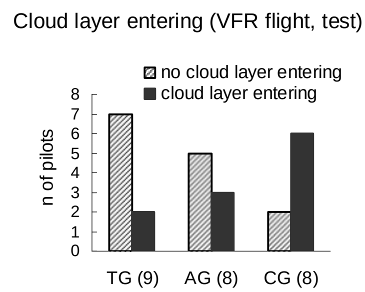


Figure 2. *Cloud layer entering during the VFR flight of the test phase for pilots of the training- (TG), awareness- (AG) and control group (CG)*

Heart Rate

Concerning the instruction flight at the beginning of the testing day, the results of the two-factorial ANOVA for repeated measures with the between factor testing group exhibit a strong main effect of the section of measurement [F(15.2,319.4) = 21.4, p = .000]. (There are neither differences between the testing groups nor is there an interaction.). As illustrated in Figure 3, the different tasks within the flight are clearly reflected in the heart rate (beats per minute, deviation from baseline). In average, the heart rate is always above the baseline. The least stressful sections are about between 90 seconds after the takeoff and 30 seconds before the first flight maneuver (role). The first strong increase of the heart rate occurs before the takeoff; descriptively the beginning of the ascent can be observed 30 sec. before the takeoff (TO), statistically (Tukey HSD post hoc tests, p < .05) it becomes significant 10 sec. before TO. When the aircraft is safely airborne, the heart rate decreases quickly within 30 seconds; the whole decrease takes about 90 sec. The flight maneuvers aileron role and looping are also reflected in the heart rate. Already 80 sec. before the landing (touchdown), there is a strong increase in the heart rate which reaches a maximum between the range of 10 sec. before and 10 sec. after the touchdown, followed by a rapid decrease within 20 seconds.

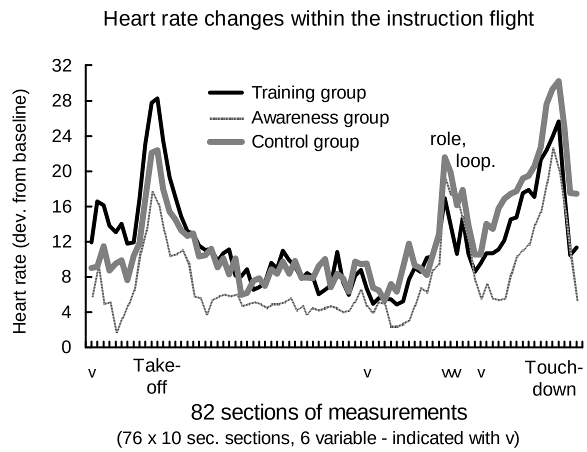


Figure 3. *Heart rate changes (beats per minute – deviation from baseline, means) separate for the three testing groups (TG: n = 8, AG: n = 8, CG: n = 8)*

Concerning the heart rate, no differences occur between the three testing groups during the flight profiles of the first two phases in the simulator.

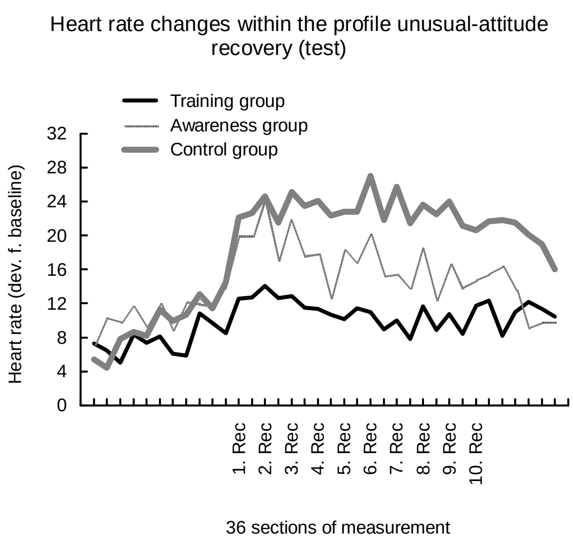


Figure 4. *Changes in the heart rate (beats per minute – deviation from baseline, means) during the test profile unusual attitude recovery (ten recoveries) for the three testing groups (training group: n = 9, awareness group: n = 7, and control group: n = 7); each recovery exercise takes about 13 seconds, the whole profile about 12 minutes.*

Within the test profile unusual-attitude recoveries, there is a clear interaction between the section of measurement and the testing group [F(17.7, 176.7) = 2.4, p = .002, Tukey HSD post hoc tests] in addition to the main effect section of measurement [F(8.8, 176.7) = 12.2, p = .000, Figure 4]. While there are no group differences at the beginning of the profile (before flying the ten recoveries), the increase of the heart rate is much higher in the control group than in the training group.

EEG – Unusual-attitude recoveries (test phase)

As analyses illustrated no differences between the three testing groups and for some calculations group sizes were too low, the factor testing group has not been involved in the following calculations. In a first step the absolute power of the EEG during waiting with closed eyes for the command to recover from an unexpected attitude (10 x 8 seconds, closed eyes), was compared with the EEG during the resting measurements before and after the test phase (each two minutes, eyes closed).

Table 4. *Average power (µV-Square) in the alpha- and beta band during the resting measurement before the test phase (RM5, 2 min., closed eyes), the time while waiting with closed eyes for the command to recover within the test profile unusual-attitude recovery (Bef. Rec., 10 x 8 sec.) and the resting measurement after the test phase (RM6), and the results of the ANOVAs*

ALPHA	RM5 (M)	Bef Rec. (M)	RM6 (M)	n	ANOVA	p-value
F3 - FC´3	2.6	2.1	3.1	17	F(2.0,32.0)=2.9	.070
F4 - FC´4	3.0	2.4	3.6	18	F(1.3,21.6)=5.7	.019
FC3 - PC3	16.5	7.9	16.5	18	F(1.4,23.1)=10.1	.002
FC4 - PC4	17.4	8.0	19.5	19	F(1.1,20.6)=9.0	.005
C3 - P3	23.7	14.5	23.0	19	F(1.7,30.2)=9.2	.001
C4 - P4	20.6	12.4	20.5	19	F(1.6,29.5)=9.1	.002
P´3 - O1	38.4	40.9	44.1	20	F(1.3,23.8)=0.5	.517
P´4 - O2	40.9	39.1	42.8	19	F(1.3,23.4)=0.7	.451

BETA	RM5 (M)	Bef Rec. (M)	RM6 (M)	n	ANOVA	p-value
F3 - FC´3	1.3	1.8	1.6	14	F(2.0,26.0)=1.2	.326
F4 - FC´4	1.2	1.5	1.3	16	F(1.3,19,.6)=1.0	.342
FC3 - PC3	4.3	3.6	4.5	15	F(2.0,28.0)=3.4	.049
FC4 - PC4	4.1	3.4	4.2	17	F(1.4,23.1)=2.7	.103
C3 - P3	4.1	3.6	4.2	16	F(2.0,30.0)=1.3	.291
C4 - P4	3.8	3.5	3.6	17	F(1.9,30.6)=0.5	.584
P´3 - O1	5.2	5.3	5.3	19	F(1.7,30.6)=0.1	.938
P´4 - O2	5.1	5.4	5.1	18	F(1.2,21.1)=0.3	.615

The results demonstrate no changes in the absolute power of the alpha band at the parieto-occipital positions P´3-O1 and P´4-O2. But concerning all other measurement positions (frontal to parietal), the alpha occurring during anticipating the recovery exercises is clearly decreased compared to a resting measurement. For the beta band, a low decrease at FC3-PC3 could be detected (Table 4).

In a second step, the periods before recovering (10 x 8 seconds, closed eyes) and during recovering (10 x 5 seconds after controls have been handed over from the instructor pilot to the participant in the simulator, eyes opened) were compared. As expected, there are of course very big decreases in the alpha band – especially over posterior regions, but at the two anterior channels, there are no changes in the alpha band. Concerning the beta band, there is a significant increase of power at F3-FC´3 and decreases at posterior regions.

Table 5. *Average power values (µV-Square) in the alpha- and beta band while waiting with closed eyes for the command to recover (Bef. Rec., 10 x 8 sec.) and while recovering (Rec. 10 x 5 sec. after controls have been taken over), and the results of the T tests*

ALPHA	Bef. Rec. (M)	Rec (M)	Diff	df	t	p-value
F3 - FC´3	1.8	1.9	.2	13	.6	.548
F4 - FC´4	2.0	1.7	-.3	13	-1.2	.259
FC3 - PC3	8.0	2.6	-5.5	14	-2.5	.026
FC4 - PC4	7.5	2.3	-5.2	14	-2.3	.037
C3 - P3	11.3	2.0	-9.4	15	-2.8	.013
C4 - P4	1.3	2.1	-8.2	15	-2.5	.026
P´3 - O1	35.8	3.2	-32.5	16	-3.1	.007
P´4 - O2	36.1	3.5	-32.7	15	-2.9	.011

BETA	Bef. Rec. (M)	Rec (M)	Diff	df	t	p-value
F3 - FC´3	1.2	1.7	**.5**	11	2.9	**.015**
F4 - FC´4	1.3	1.6	.3	13	1.3	.220
FC3 - PC3	2.8	2.6	-.2	12	-.8	.423
FC4 - PC4	3.5	2.3	-1.2	13	-2.0	.071
C3 - P3	3.5	2.4	-1.1	14	-2.9	.012
C4 - P4	3.5	2.1	-1.4	14	-2.2	.042
P´3 - O1	5.2	3.0	-2.2	16	-3.2	.005
P´4 - O2	5.3	3.0	-2.3	14	-3.9	.002

Discussion

The results of the flight performance data confirm positive training effects, especially for the test profile "VFR flight at minimal weather conditions, mountains". Pilots with a training behave less risk prone, whereas pilots without any kind of orientation training do often not turn back at an appropriate moment. They enter the cloud layer more frequently and lose orientation, which finally can lead to a crash into the mountain or into the ground by trying to stay under the cloud layer without realizing that the mountains are in clouds. This happened despite the fact that the pilots had a map (including the geographical data of the region etc.), that they had flown the route already under conditions of good visibility (instruction flight), and that they always had the possibility to get weather information from the "air traffic controller" (i.e. form the instructor pilot at the external work station of the simulator). As many accident reports, this fact highlights the problem of deteriorating visibility conditions: Some VFR pilots do not recognize the ensuing danger which can lead to fatal crashes, even in regions well known to the pilots.

The simulation scenarios are of high impact for the pilots, as could be demonstrated by the changes in the heart rate. As example the data of the instruction flight have been presented. The clear increases caused by the takeoff and the landing procedure are similar to the published results concerning changes during flight (e.g. Hankins & Wilson, 1998; Wilson, 2002). Veltman (2002) compared psychophysiological reactions during simulator and real flight and could confirm similar results for heart rate, heart rate variability, and respiratory frequency.

Our analysis of the heart rate within the test profile unusual-attitude recovery demonstrates the expected increase of heart rate due to increasing mental workload. Additionally, the results illustrate a significant lower increase of the heart rate for pilots having received an unusual-attitude recovery training. As a conclusion, the effects of the evaluated training program can be described as increasing flight performance together with reducing stress in demanding flight situations. First EEG results show changes in the alpha- and the beta bands due to changing strain in the simulator.

To sum up, this study makes a contribution to basic research by analyzing psychophysiological changes as well as to applied science by emphasizing the importance and effectiveness of orientation training programs for VFR pilots.

Acknowledgments

The research was granted by the Austrian Industrial Research Promotion Fund (FFF).

References

AMST Systemtechnik. Flight simulators and technical parameters. Retrieved February 10, 2005, from *http://www.amst.co.at.*

Endsley, M. R. (1995). Measurement of Situation Awareness in Dynamic Systems. *Human Factors, 37* (1), 65-84.

Endsley, M. R. (2000). *Theoretical Underpinning of Situation Awareness: A Critical Review.* In Endsley, M. R. & Garland, D. J. (Eds.), Situation Awareness Analysis and Measurement (pp 3-32). New Jersey: Lawrence Erlbaum Associates.

Goh, J. & Wiegmann, D. A. (2001). Visual Flight Rules Flight into Instrument Meteorological Conditions: An Empirical Investigation of the Possible Causes. *International Journal of Aviation Psychology, 11*(4), 359-379.

Hankins, Th. C. & Wilson, G. F. (1998). A Comparison of Heart Rate, Eye Activity, EEG and Subjective Measures of Pilot Mental Workload During Flight. *Aviation, Space, and Environmental Medicine, 69*(4), 360-367.

Haug, T. (2003). *AMST SD Private-Pilot Flight Training. Instructor Pilot Checklist*. Ranshofen: Unpublished Training manual.

Hoffmann, J. (1993). *Vorhersage und Erkenntnis. [Prediction and Recognition]* Göttingen: Hogrefe.

Kallus, K. W. & Tropper, K. (2004). Evaluation of a Spatial Disorientation Simulator Training for Jet pilots. *International Journal of Applied Aviation Studies, 4*(1), 45-55.

Veltman, J. A. (2002). A Comparative Study of Psychophysiological Reactions During Simulator and Real Flight. *International Journal of Aviation Psychology, 12*(1), 33-48.

Véronneau, S. J. H. & Evans, R. H. (2004). *Spatial Disorientation Mishap Classification, Data, and Investigation.* In F. H. Previc & W. R. Erocline (Eds.), Spatial Disorientation in Aviation (pp. 197-241). Virginia: American Institute of Aeronautics and Astronautics.

Wickens, C. D. (2002). Situation Awareness and Workload in Aviation. *Current directions in Psychological Science, 11*(4), 128-133.

Wilson, G. F. (2002). An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *International Journal of Aviation Psychology, 12*(1), 3-18.

IMPLEMENTING ELECTRONIC FLIGHT DATA IN AIRPORT TRAFFIC CONTROL TOWERS

Todd R. Truitt, Ph.D.
FAA, William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

The Federal Aviation Administration (FAA) is investigating the potential effects of implementing electronic flight data systems (EFDSs) at Airport Traffic Control Towers (ATCTs). I use existing task analyses, published literature, and recent field observation data to determine the basic functionality of flight progress strips (FPSs) in the ATCT. I identify gaps in the research and formed a general set of principles to guide the design of an EFDS prototype. Given the proper application of principles for design and automation, the EFDS should maintain some of the basic functionality and benefits of the FPSs, reduce workload related to flight data entry, tracking and sharing, and provide new features that will enhance controller performance and encourage use. I present possible risks and outcomes that are likely to accompany an EFDS in FAA ATCTs.

Background

Airport operations logged by the 449 Federal Aviation Administration (FAA) airport traffic control towers (ATCTs) are projected to increase from 62.7 million in 2003, to 70 million in 2007 (FAA, 2004a). In anticipation of the increase in air traffic, the FAA is investigating the potential effects of implementing an electronic flight data system (EFDS) in ATCTs. One primary interest is how to preserve the current benefits of paper flight progress strips (FPSs) while enhancing the performance of air traffic controllers and the National Airspace System (NAS). To do so, we must understand the similarities and differences among ATCTs as well as all of the tasks involving FPSs, flight data, and the communication of information among air traffic controllers. Researchers can contribute to the success of an EFDS if they address some major gaps in the existing research and address long-standing organizational norms during the design process.

In general, the controller positions in an ATCT include flight data (FD), clearance delivery (CD), ground control (GC), and local control (LC). ATCTs often combine the FD and CD positions during periods of lower taskload, and some ATCTs may staff positions in addition to those just mentioned (FAA, 2004b). Each controller position has a general set of duties. Typically, the FD/CD position enters flight plans and flight plan amendments into the computer, distributes flight data, issues initial long-range clearances, enters and updates the automatic terminal information service (ATIS) information, and coordinates clearances with air route traffic control centers. The GC position provides aircraft and vehicle taxi instructions to and from the airport movement area and the ramp and gate area, coordinates crossing or use of active runways, and determines the departure sequence. The LC position provides departure and arrival sequencing and spacing by issuing clearances to all aircraft in the airport traffic area and all aircraft and vehicles on the active runways. Both the GC and LC positions may be required to coordinate among multiple other LC and GC positions.

Among the 449 ATCTs in the United States, each provides a particular type of service including visual flight rules only, non-radar, or radar approach control. Within each ATCT, there are different types of equipment, specific controller positions, and duties that vary by facility. Each ATCT typically has its own facility directive that provides a set of supplemental standard operating procedures to address local idiosyncrasies.

How Controllers Use FPSs in the ATCT

Even though there is substantial variability among ATCTs, the use of FPSs is relatively ubiquitous. In addition to FPSs, controllers use other sources of information along with tools for communication, coordination, information organization, and decision making. However, one of the arguably central tools used in the ATCT along with the radio, is the FPS (Bruce, 1996; FAA, 2004c). The use of FPS has a long history, and since their inception in the 1930's and 1940's, very little has changed. Over time, the FAA has rooted the FPS through training regimens, handbooks, standard operating procedures, and facility directives. There is currently a significant amount of pressure exerted upon controllers to use FPSs (Durso & Manning, 2002).

Because the use of FPSs and the information they contain has become an integral part of the ATCT task, it is important to understand how controllers use FPSs in the ATCT domain and how the FPSs aid in the flow of information. Acknowledging differences among ATCTs, the general flow of information for departure aircraft is from FD/CD to GC to LC to terminal radar control (TRACON). For arrival aircraft, the information moves in the opposite direction from the TRACON to LC to GC. The type of information that controllers pass among each another varies too depending on the phase of an aircraft's flight (e.g., arrival, departure, or over flight).

The differences among ATCTs and individual controllers also reflects in the functions that FPSs serve. While controllers amend the FPSs using a standard set of symbols in accordance with the 7110.65P (FAA, 2004c) and a few unique markings as published in their own facility directive, there are also individual preferences for FPS use. While the individual needs of ATCTs and controllers are important, it is not yet necessary to understand how every one conducts operations in particular. We must first collect empirical evidence regarding the critical functions of FPSs and how to best support these functions with an EFDS.

It is clear that controllers use the FPSs and their associated markings for a number of purposes. A number of researchers have examined the particular functions of FPSs, whereas others have examined the higher-level cognitive processes that controllers support with FPSs. These researchers have shown that across various ATC domains controllers use FPSs for workload management (Durso & Manning, 2002; Gronlund, Dougherty, Durso, Canning & Mills, 2001; Dattel, Johnson, Durso, Hackworth & Manning, 2005), memory aids (Buisson & Jestin, 2001; Cardosi, 1999; Durso & Manning; Gronlund et al.; Hopkin, 1988; Dattel et al.; Pavet, 2001; Stein & Bailey, 1994; Zingale, Gromelski, Ahmed, & Stein, 1993; Zingale, Gromelski, & Stein, 1992), facilitating communication and coordination (Berndtsson & Normark, 1999; Buisson & Jestin; Durso & Manning; Gronlund et al.; Dattel et al.; Pavet), cognitive information organization (Durso & Manning; Dattel et al.), and planning (Cardosi; Dattel et al.; Gronlund et al.; Pavet; Zingale et al.). However, researchers have debated the necessity of FPSs and their use. A primary debate has centered on whether or not the FPSs provide any real benefit to memory, and ultimately, performance.

While researchers have conducted a number of studies in the en route domain, the debate between the Interaction and Cognitive Resource hypotheses (for a brief review, see Vortac, et al., 1996) has not surfaced in the ATCT domain until now. In fact, researchers conducted only a few controlled studies to understand what controllers are doing in the ATCT and how they are doing it. Bruce (1996) conducted a study that focused on the physical performance of controllers in the ATCT and provided valuable information about what controllers did while working. For example, her data showed that controllers most often manipulated FPSs, microphones, and writing pens. Along with their human abilities, these are the controllers' primary tools. Bruce also showed that GCs spent about one-half of their time directly observing traffic out of the window, whereas LCs spent only about one-third of their time looking outside. Incidentally, the LC's time observing traffic doubled when radar data were available in the ATCT.

Ammerman, Becker, Bergen, et al. (1987), Ammerman, Becker, Jones, et al. (1987), and Alexander, et al. (1989) published a comprehensive set of task analyses of ATCT activity, which are still relevant today. Alexander et al. examined the baseline, or current activity, of ATCTs, while Ammerman, Becker, Bergen, et al. explored the future concept of the Tower Control Computer Complex (TCCC) envisioned within the Advanced Automation System concept. As the name implied, the TCCC was to rely more on computer power, shared information, and automation and rely less on pen and paper. Some of the concepts envisioned for the TCCC like Airport Surface Detection Equipment (ASDE) have materialized while others, like reconfigurable tower position consoles at each controller position, have not. Despite the current state of affairs, these task analyses are still valuable today in that they provide, among other things, compositional graphs that show the logical flow of operational tasks, information requirements, and necessary cognitive/sensory attributes.

Researchers have conducted numerous other studies as well, but these studies have lacked the data required to consider hypotheses regarding the cognitive effects of an EFDS in FAA ATCTs. Nevertheless, this past research is very helpful in providing insights into risks and benefits of an EFDS. For example, Chistophe Mertz and his co-authors present an array of interface usability research that provides many valuable lessons on the use of touch screens in air traffic control (e.g., Mertz, Chatty, Vinot, 2000a, 2000b; Mertz & Lecoanet, 1996; Mertz & Vinot, 1999). Doble and Hansman (2003) examined the concept of using pocket computers to replace FPSs; a concept that Buisson and Jestin (2001) also explored. These authors present significant insight into the advantages and limitations of using pocket computers as FPS replacements.

Only recently have researchers collected data specifically on controllers' FPS activity in the ATCT. Dattel et al. (2005) used subject matter expert observers to record controllers' FPS marking and handling behavior during live operations. Their observations included the three primary control positions (FD/CD, GC, LC) at 10 ATCTs located across the United States. The ATCTs were of various sizes and handled differing levels and complexity of traffic. The authors examined both the frequency and the importance of FPS marking by controller position and facility size. In addition, they followed the observation sessions with directed interviews and questionnaires to gain insight about the perceived psychological benefits of FPSs including communications, memory, organization, situation awareness, and workload. Dattel et al. found that each controller position used the FPSs for different reasons, and these uses did not depend on facility size. Controllers at the FD/CD position reported that FPS activity benefited communication, workload, and memory. FD/CD used marking primarily for the benefit of others. Controllers at the GC position reported that FPS activity supported all five psychological functions. Controllers at the LC position reported FPS benefits for memory, organization, and situation awareness. However, controllers at both the GC and LC positions believed that the primary benefits of FPS were associated with memory and situation awareness. Researchers have yet to determine whether any of these reported benefits are actual or just perceived, and if they are real, the size and duration of any effect on controllers' performance.

An Alternative to FPSs

Replacing the FPSs used in the ATCT with an EFDS would require new hardware, procedures, and automation that relieve the controller of workload arising from non-essential, "housekeeping" tasks while improving performance. Performance could benefit simply by reducing the workload associated with FPSs, but properly designed interfaces and automation could elevate performance beyond that which controllers might obtain only by addressing workload. A feasible EFDS in the ATCT should integrate the controller's perceptual abilities with improvements in navigation, radar, and automation including weather detection and traffic alerting systems (Ammerman, Becker, Bergen, et al., 1987). The EFDS should provide the same proven critical benefits as FPS while eliminating outdated uses such as recording of some clearances to establish a legal record. The EFDS, resting on the concept of System Wide Information Management (SWIM) (FAA, 2004d) will provide new functionality through automation, especially in terms of information sharing. Such new functionality should make some current tasks easier and provide controllers with the ability to perform actions that they could not perform with FPSs.

There are a number of features that an EFDS could provide in the ATCT. The ability to display and input flight data from a single interface opens many possibilities, but the ability to share information among various systems is what will make an EFDS especially useful. Information will be able to move between a flight data element and any other component of the primary system. Two-way information updates provide easy access and sharing of flight data such as clearance amendments, predicted runway/taxiway incursions, aircraft location on a taxiway, posting and updating expected departure clearance times, alerts for traffic flow restrictions, and wake turbulence warnings. An EFDS allows for the elements of one or more situation displays to be linked so that items of interest can be emphasized and identified simultaneously for categorization. Electronic flight data elements can appear only when controllers need them the most and still preserve the ability to access all information about any flight at any time. An EFDS would provide an interface for digital communications such as controller-pilot data link communications (CPDLC). CPDLC via the EFDS interface would allow the controller to provide flight information services (e.g., pilot reports, weather reports, maps, approach plates, etc.), pre-departure clearances, full taxi instructions including gate information and visual depiction of taxi route, digital ATIS (D-ATIS), and even landing and takeoff clearances. An EFDS also allows for simplified data input such as recording certain clearances or updating an ATIS code with simple motions or gestures while preserving the ability to make freehand notation. Moreover, all data entries on an EFDS are shared and become available to other controllers as necessary. Researchers have already designed automation tools that could potentially be integrated with an EFDS under the SWIM concept. Such tools may provide assistance with taxi sequencing (e.g., Departure Planner Decision Aid, Anagnostakis, et al., 2000) changing runway configuration (e.g., Surface Management System, Atkins & Brinton, 2002), and digital watermarking (e.g., Hering, Hagmüller & Kubin, 2003; Prinz, Sajatovic, & Hering, 2004).

The potential advantages of an EFDS are numerous. An EFDS would eliminate workload associated with placing FPSs in holders, distributing FPSs, and handling multiple FPSs for a single flight. Controllers may increase the time they spend looking out the window of the tower cab and directly observing the traffic situation. Controllers also may increase their awareness of others controllers' actions through the use of both distributed displays that share flight data elements and through the use of shared displays (Mertz & Lecoanet, 1996). Flight data activity that is currently tallied by time-consuming, manual processes could be automatically tracked on an EFDS to allow for automatic traffic counts and the recording of timing information and clearances. An EFDS simplifies the act of passing flight data among controller positions within the ATCT and between the ATCT and TRACON. Electronic flight data allows controllers to pass information virtually rather than having to move away from their control position and physically transfer a FPS. An EFDS even creates the potential for saving money budgeted for the purchase of paper FPSs, FPS holders, and the maintenance of the thermal printers.

The potential disadvantages of an EFDS are not as obvious as the advantages. I have already discussed the need for researchers to learn about the effects that any new system will have on users. If the EFDS does affect controller performance, the extent and direction of change will depend in part on the design of the EFDS and on how the FAA trains controllers to use it. Even if an initial decrement in performance does occur, controllers may be able to overcome changes to their task rather quickly. Unfortunately, there currently aren't any data on the ATCT domain to inform us about the effects of changing the format of flight data information or changing the way that controllers interact with flight data. Previous data suggests that although the new EFDS will not eliminate physical interaction with flight data, it may change the frequency and types of interactions that controllers perform. Such a change in behavior may have positive or negative effects upon controllers' performance (e.g., Vortac et al., 1996) memory (e.g., Hopkin, 1988; Stein & Bailey, 1994; Zingale, Gromelski & Stein, 1992), or situation awareness (Endsley & Rodgers. 1996; Hopkin, 1995). However, these are empirical questions that researchers must still answer within the ATCT domain.

Another potential disadvantage of an EFDS is that a pen- or gesture-based system may be more difficult to use than paper FPSs, especially at first (Mertz & Vinot, 1999; Mertz, Chatty, & Vinot, 2000b). Data entry will also become more critical as more information is shared with more people (Della Rocco, Manning, & Wing, 1990). We can't forget that this flight data is being used for safety critical functions. Data entry errors could potentially result in other, more serious unwanted outcomes. EFDS designers should make data entry as easy as possible and methods for identifying and correcting errors are needed. The transition from FPSs to an EFDS may also impact the controller selection and training process rendering them less useful and in need of modification (Della Rocco et al.).

The FAA recently implemented a policy establishing that no new displays occupy the ATCT except by an explicit waiver process. This "no new glass" policy arose from the numerous systems that have already been deployed in the ATCT. Not only have these new systems taken up precious space inside the tower cab, they also operate independently of one another. In other words, the FAA has filled the ATCT with a multitude of non-integrated systems creating a crowding of the physical space, increased maintenance costs, and the inability of systems to cooperate with one another.

Given the FAA's "no new glass" policy and the various levels of traffic and technology at the 449 ATCTs in the United States, it is very likely that different EFDSs may have to be developed for different types of ATCTs. For example, ATCTs that have ASDE or other types of surface radar displays may be able to take advantage of an existing data source by integrating the flight data with it. The suggestion of integrating flight data with surface radar data is a viable one. Such an approach has already begun at Nav Canada. Airports without ASDE could still take advantage of an EFDS, but the optimal presentation of flight data may require a different form. To take full advantage of electronic flight data, FAA researchers must consider deploying alternative perceptual-spatial displays that don't rely on ASDE. There is one thing that we know about ATCTs; there is a great deal of variation and one solution will not fit well for all.

Whatever form any new features take, they must be reliable, provide valid information, and have a wide and demonstrable effect before controllers are likely to accept them. The new features that an EFDS would enable should also provide some incentive for controllers to overcome the well-entrenched FPS and to adopt the new EFDS. By providing an irresistible alternative to FPS, I hope overcome the organizational norms that have made FPSs a well-entrenched tool in the ATCT domain.

Making the Transition

Beyond providing new tools for controllers, researchers and system designers must also get participation from controllers and controller union representatives during the entire research and development process to aid in overcoming the organizational norms that embody FPSs. Controllers should serve as subject matter experts to help researchers understand the ATCT domain and to provide insight on interface design and functionality. By involving controllers throughout the entire process, the FAA can get the support that will be needed when change is upon the controllers. Furthermore, controllers will have a stake in the process and be anticipating the change knowing that the transition to an EFDS will be worthwhile because researchers and system designers have already considered their actual job requirements

Summary and Conclusion

Having the support of controllers is a necessary condition, but not sufficient to ensure the success of an EFDS. Researchers also need to learn more about the psychology of FPSs. As previously mentioned, there is very little data concerning how controllers in the ATCT perceive and gather flight data, but the ATCT domain poses some familiar questions. The Interaction and Cognitive Resource hypotheses become relevant again. It is appropriate and necessary to ask these same questions again because the task of controllers in the ATCT is quite different than that of controllers in the en route environment. Our knowledge of how controllers use FPSs in the en route domain does not allow us to fully understand other domains. During the development of an EFDS for the ATCT, we must know if changes to the presentation of flight data in an EFDS will affect the controllers' ability to find or use that information. We must know if the controllers' ability to find and use flight data will be affected by the way they physically interact with the system. Researchers need to employ various part-task or low-fidelity simulations to understand basic cognitive functions, but they must also perform high fidelity, human-in-the-loop simulations to test the concepts they create. With the support of empirical data and proper system design, the FAA will be able to capitalize on the benefits of an EFDS and mitigate the associated risks.

References

Alexander, J. R., Alley, V. L., Ammerman, H. L., Fairhurst, W. S., Hostetler, C. M., Jones, G. W., & Rainey, C. L. (1989). *FAA air traffic control operations concepts. Volume VII: ATCT tower controllers*. (FAA Report No. DOT/FAA/AP-87-01, Vol. 7). Washington, DC: Federal Aviation Administration.

Ammerman, H. L., Becker, E. S., Bergen, L. J., Claussen, C. A., Davies, D. K., Inman, E. E., & Jones, G. W. (1987). *FAA air traffic control operations concepts. Volume V: ATCT/TCCC Tower Controllers*. (FAA Report No. DOT/FAA/AP-87-1, Vol. 5). Washington, DC: Federal Aviation Administration.

Ammerman, H. L., Becker, E. S., Jones, G. W., Tobey, W. K., & Phillips, M. D. (1987). *FAA air traffic control concepts. Volume I: ATC background and analysis methodology*. (FAA Report No. DOT/FAA/AP-87-01, Vol. 1). Washington, DC: Federal Aviation Admistration.

Anagnostakis, I., Idris, H., Hall, W., Clarke, J-P., Feron, E., Odoni, A., & Hansman, R. J. (2000, November). *A conceptual design for a departure planner decision aid*. Presented at the 3rd International Seminar on ATM R&D. Napoli, Italy.

Atkins, S., & Brinton, C. (2002). Concept description and development plan for the surface management system. *Journal of Air Traffic Control, 44(1).*

Berndtsson, J. & Normark, M. (1999). The coordinative functions of flight strips. *Proceedings of the International ACM Siggroup Conference on Supporting Group Work* (pp. 101-110). New York: ACM Press.

Bruce, D. S. (1996). *Physical performance criteria for air traffic control tower specialists*. Federal Aviation Administration: Washington, DC.

Buisson, M., & Jestin, Y. (2001). Design issues in distributed interaction supporting tools: Mobile devices in an ATC working position. *Proceedings of Mobile HCI 2001: Third International Workshop on Human Computer Interaction with Mobile Device.*,Lille, France.

Cardosi, K. M. (1999). *Human factors for air traffic control specialists: A user's manual for your brain* (DOT/VNTSC-FAA-99-6). Cambridge, MA: US Department of Transportation, John A. Volpe National Transportation Systems Center.

Dattel, A. R., Johnson, B. R., Durso, F. T., Hackworth, C., & Manning, C. A. (2005). Flight progress strips in towers: Frequency index and perceived psychological benefits. *Proceedings of the 13th International Symposium on Aviation Psychology*. Oklahoma City, OK.

Della Rocco, P. S., Manning, C. A., & Wing, H. (1990). *Selection of air traffic controllers for automated systems: Applications from current research* (DOT/FAA/AM-90/13). Washington, DC: Office of Aviation Medicine.

Doble, N. A., & Hansman, R. J. (2003). *Design and evaluation of a portable electronic flight progress strip system* (Report No. ICAT-2003-4). Cambridge, MA: MIT International Center for Air Transportation.

Durso, F. T., & Manning, C. A. (2002). Spinning paper into glass: transforming flight progress strips. *Human Factors and Aerospace Safety, 2(1)*, 1-31.

Endsley, M. R., & Rodgers, M. D. (1996). Attention distribution and situation awareness in air traffic control. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 82-85).

Federal Aviation Administration (2004a). *Administrator's Fact Book*. Washington, DC.

Federal Aviation Administration (2004b). *Facility Operation and Administration* (7210.3T). Washington, DC.

Federal Aviation Administration (2004c). *Air Traffic Control* (7110.65P). Washington, DC.

Federal Aviation Administration (2004d). *System-wide information management (SWIM) architecture and requirements* (CNS-ATM Task 17 Final Report TR04013). Washington, DC.

Gronlund, S. D., Dougherty, M. R. P., Durso, F. T., Canning, J. M., & Mills, S. H. (2001). *Planning in air traffic control* (DOT/FAA/AM-01/16). Washington, DC: Office of Aerospace Medicine.

Hering, H., Hagmüller, M., & Kubin, G. (2003). Safety and security increase for air traffic management through unnoticeable watermark aircraft identification tag transmitted with the VHF voice communication. *Proceedings of the 22nd Digital Avionics Systems Conference* (pp. 4.E.2-1 – 4.E.2-10). Indianapolis, IN: IEEE.

Hopkin, V. D. (1988). Air traffic control. In E. L. Wiener and D. C. Nagel (Eds.), *Human factors in aviation* (pp. 639-663). San Diego: Academic Press.

Hopkin, V. D. (1995). Situational awareness in air traffic control. In R. D. Gilson, D. J. Garland, & J. M. Koonce (Eds.), *Situational awareness in complex systems: Proceedings of a CAHFA conference* (pp. 171-178). Daytona Beach: Embry Riddle Aeronautical University Press.

Mertz, C., & Lecoanet, P. (1996). GRIGRI: Gesture recognition on interactive graphical radar image. *Proceedings of the Gesture Workshop'96*. York, United Kingdom.

Mertz, C., & Vinot, J. (1999). Touch input screens and animations: More efficient and humanized computer interaction for ATC(O). *Proceedings of the 10th International Symposium on Aviation Psychology* (615-621). Columbus, OH.

Mertz, C., Chatty, S., & Vinot, J. (2000a). Pushing the limits of ATC user interface design beyond SandM interaction: the DigiStrips experience. *Proceedings of the 3rd USA/Europe Air Traffic Management R&D Seminar*. Napoli, Italy.

Mertz, C., Chatty, S., & Vinot, J. (2000b). The influence of design techniques on user interfaces: The DigiStrips experiment for air traffic control. *Proceedings of HCI-Aero 2000 International Conference on Human-Computer Interaction in Aeronautics*. Toulouse, France.

Pavet, D. (2001). Use of paper strips by tower air traffic controllers and promises offered by new design techniques on user interface. *Proceedings of the 4th USA/Europe Air Traffic Management R&D Seminar*. Santa Fe, NM.

Prinz, J., Sajatovic, M. & Hering, H. (2004). S^2EV-Analog voice with enhanced safety and security. *Proceedings of the 49th Annual ATCA Conference*. Washington, DC.

Stein, E. S., & Bailey, J. (1994). *The controller memory guide: Concepts from the field* (DOT/FAA/CT-TN94/28). Atlantic City, NJ: US Department of Transportation, Federal Aviation Administration.

Vortac, O. U., Barile, A. L., Albright, C. A., Truitt, T. R., Manning, C. A., & Bain, D. (1996). Automation of flight data in air traffic control. In D. Herrmann, M. Johnson, C. McEvoy, C. Hertzog, & P. Hertel (Eds.), *Basic and applied memory: Research on practical aspects of memory*, *2*. Hillsdale, NJ: Erlbaum.

Zingale, C., Gromelski, S., Ahmed, S. B., & Stein, E. S. (1993). *Influence of individual experience and flight strips on air traffic controller memory/situational awareness* (DOT/FAA/CT-TN93/31). Atlantic City, NJ: US Department of Transportation, Federal Aviation Administration.

Zingale, C., Gromelski, S., & Stein, E. S. (1992). *Preliminary studies of planning and flight strip use as air traffic controller memory aids* (DOT/FAA/CT-TN92/22). Atlantic City, NJ: US Department of Transportation, Federal Aviation Administration.

EID OF A PILOT SUPPORT SYSTEM FOR AIRBORNE SEPARATION ASSURANCE

Stijn B.J. Van Dam, **An L.M. Abeloos**, **Max Mulder** and **René (M. M.) van Paassen**

Delft University of Technology, Faculty of Aerospace Engineering, Control and Simulation Division
Delft, The Netherlands

In a flexible airspace environment the pilot disposes of an increased amount of travel opportunities. At the same time the airspace traffic situation becomes more complex and the aircraft separation assurance task is shifted towards the cockpit. The design paradigm of Ecological Interface Design is applied to support the pilot with the airborne planning of efficient trajectory paths that maintain spatial separation from other traffic. The desired pilot behavior is achieved by visualizing travel-relevant airspace affordances in terms of realistic aircraft locomotion. As a result, a novel interface the ”state vector envelope” presents safe and efficient travel opportunities in a state vector field. The concept has been evaluated through on-line simulations of a number of basic conflict situations.

Introduction

In traditional airspace environment, capacity problems are expected in the near future due to growing air traffic, hereby causing a higher workload for Air Traffic Controllers. New concepts for Air Traffic Management such as Free Flight permit a flexible use of airspace with airborne determination of User-Preferred Trajectories or UPT's (ACM, 2002) which allow direct routing and cruise climb tasks. This flexible use is expected to increase airspace capacity and improve congestion problems.
The separation task is shifted from the air traffic controller towards the pilot and it is expected that the latter needs to be assisted in this task. The question is whether current systems always fully exploit numerous travel opportunities offered by a more flexible and complex airspace. This extended pilot navigation task of trajectory planning, including separation, needs to be supported by a more general airborne trajectory planning system.

New technologies have already made it possible to assure spatial separation from other aircraft in the cockpit with the so called Airborne Separation Assurance Systems ACM, 2002. These systems predict when spatial separation is going to be lost (conflict detection), communicate this event to the pilot and provide and suggest resolutions (conflict resolution).

ASAS systems, as for example developed by the Dutch Aerospace Laboratory NLR (Hoekstra, 2001) have proved to offer the pilot a safe and effective conflict detection and resolution with speed and heading markers. Unfortunately, the system can not prevent that the aircraft resolution maneuver resolves one conflict, but triggers another. In the same way, it can not prevent the occurrence of very dangerous short term conflict situations due to trajectory changes like leveling off or turning.

A further development of the NLR system, the Predictive ASAS system or P-ASAS informs the pilot about which state changes would trigger new conflicts by the use of individual no-go state bands on the speed taper and heading scale. Each no-go zone holds for maneuvers in that state dimension. Therefore its use to prevent short term conflict situations is only applicable to aircraft maneuvers that consist of a sole heading or speed change.

Further improvements on these systems should be possible. However, in our opinion there must be a better way to support airborne trajectory planning. The P-ASAS system calculates and presents an explicit automatic solution, which disables the pilot from integrating other trajectory planning- relevant tasks with the spatial separation task. Previous research at the Delft University of Technology (Hoekstra, 2001 & van Paassen, 1999) does not aim to calculate and present an explicit automated solution, but starts from the exploration and presentation of conflict-free trajectory possibilities. Such a presentation helps the pilot to both resolve and prevent conflict situations while the freedom to consider other travel-relevant aspects into the trajectory planning task is preserved.

Besides the fact that it does not support efficient conflict resolution, the guidance tools related to the former locomotion models with instant heading change or turn maneuvers show a high sensitivity to flight speed changes. The no-go zones split up, enlarge or shrink, move from one side to another. Therefore, the research presented in this paper explores the potential of a locomotion model that incorporates the ground speed change to present efficient conflict-free trajectory travel guidance. The aircraft dynamics will be neglected due to the complexity of expressing combined heading and speed changes.

This project took a cognitive engineering approach. As mentioned before, spatial separation is not the only pilot task that needs to be performed for efficient and safe airspace travel. A workspace analysis of the airborne trajectory planning task defines a complete overview of the pilot's work domain. It reveals hierarchic relations between travel physics, planning tasks and the achievement of travel goals in terms of safety, production and efficiency. These relations are made directly visible for ecological interface design by applying a functional modeling technique based on the perception of environmental affordances (Gibson, 1979). At this stage the locomotion model is studied within the conflict geometry and dynamics. As a result a more functional or meaningful, rather than pure physical presentation of aircraft and airspace physics helps the pilot to see the travel opportunities with respect to his planning task. The interface is evaluated through an experiment with on-line simulations of a number of basic conflict situations. Conclusions and recommendations regarding the ecological interface design are given at the end of the paper.

ATP Work Domain

In a flexible use airspace environment the extended navigation task, which includes spatial separation, will be defined as airborne trajectory planning task.

Airborne Trajectory Planning (ATP) is a general concept addressing the on-board planning of a travel goal satisfying trajectory path within a flexible use airspace environment

By setting up an abstraction hierarchy table (Figure 1) for this task, travel goals, on functional purpose level, are related to the abstract key functions and to physical models of airspace and aircraft. This way a multi-level overview of the pilot's work domain is obtained. The key functions will be used to set out the planning task description.

The ATP systems' main goal is traveling through airspace. Three sub-goals are identified on a functional purpose level: safety, production and efficiency. On a abstract functional level the key functions reveal how the goals can be achieved. On the general functional level traveling and path control have to realize these key functions. On the bottom of the table the aircraft and airspace model represent the physical form of the system. For trajectory planning, the workspace is reduced to short- and middle term locomotion issues in the horizontal plane.

General air transportation key functions such as staying inside the flight envelope, assuring propulsion and lift, providing passenger comfort are not relevant for the locomotive trajectory path. Although trajectory planning is done in a 3D airspace environment, vertical maneuvers are excluded in order to focus on the horizontal space domain. In the horizontal plane an aircraft will travel towards a chosen waypoint or destination. As the planning task is applied to multiple conflict situations, the look-ahead time for conflict support ranges from short to middle term.

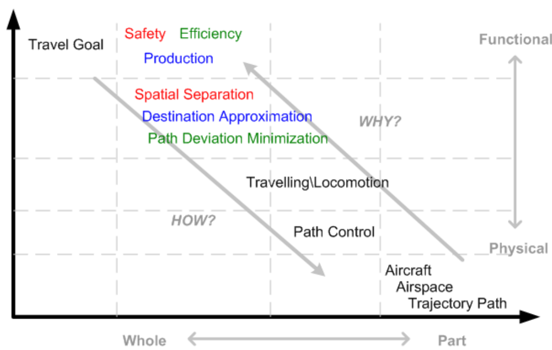


Figure 1. *Abstraction Hierarchy for Airborne Trajectory Planning system.*

Now that the domain boundaries are clearly outlined, the identified Key Functions involved are spatial separation, path deviation minimization and destination approximation. These key functions are evaluated against measurable physical criteria. These criteria are deduced from physical properties of aircraft and airspace and the relation between these properties.

Spatial Separation (safety goal). The violation space around an aircraft is defined in the separation criteria ACM, 2002. The point on the trajectory prediction, which lies within the 5 minute look-ahead time and where the spatial separation with the other aircraft is smaller than the minimal value, is called the Closest Point of Approach (CPA). In the horizontal plane the CPA distance has to be larger than 5 nautical miles.

Destination Approximation (production goal). A destination can be the destination of a flight, but also the next waypoint or the next entry or exit point on the border of another airspace. Often, spatiotemporal requirements need to be met for arriving at that point: a maximal spatial and time deviation. For this study, a simple requirement stating that the distance between aircraft and destination should always decrease in time (therefore called destination approximation), will be used.

Path Deviation Minimization (efficiency goal). The path deviation is parameterized by the maximal spatial deviation. This distance is the 3D distance between the original and alternative trajectory position point at a given time instance. After passing the closest Point of Approach, the traveler starts its recovery maneuver towards the original trajectory. At this point the deviation distance is a measure for the conflict resolution efficiency.

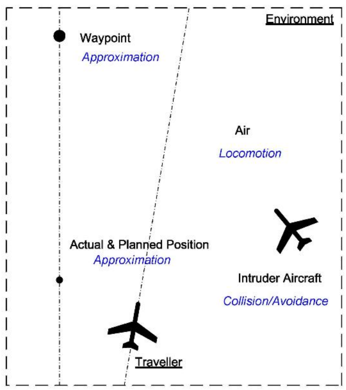


Figure 2. *Travel Space Modeling by the notion of affordances. Airspace elements afford trajectory planning relevant properties.*

Functional Modeling Based on the Perception of Affordances

The step of translating cognitive work analysis of a complex work domain into an interface design is based on a ecological interface design concept developed by Vicente & Rasmussen (Vicente and Rasmussen, 1992). Functional Modeling tries to formulate the behavior of a system relevant to achieving its ends (Lind, 1990). A paradigm of ecological psychology, the perception of affordances (Gibson, 1979), describes the human capacity to directly perceive and act upon environmental affordances, rather than the assessment of physical qualities or properties. For trajectory planning, the goal relevant affordances must be formulated or visualized in such a way that the perception of these by the pilot, directly triggers desired goal relevant or functional aircraft behavior by the pilot's steering actions. Figure 2 provides a pictorial overview. The surrounding unoccupied air provides the affordance of locomotion to the aircraft, other aircraft in the vicinity provide the affordance of collision (or the opposite, avoidance). Note that currently the listed affordances are not yet visualized adequately.

Locomotion Model

In order to assure that the perception of affordances can be fluently transformed in functional aircraft behavior, the affordances are formulated in terms of aircraft locomotion that matches flight practice. For trajectory planning in a cruise flight limited to the horizontal plane, the pilot determines its aircraft behavior by manipulation of heading and airspeed settings, while the autopilot flies on altitude hold mode. Therefore, a locomotion model should yield heading and/or speed change. In this way the model, reduced to a one or two dimensional input, is less complex and more practical than a traditional multidimensional state space presentation. As explained in the introduction, the two first locomotion models explore travel opportunities through heading changes, either instantaneous or including realistic turn dynamics. Because of their lack of conflict resolution efficiency and their sensitivity to speed changes, a third model is built which combines speed and heading changes.

Visualization of Affordances

For productive planning, the affordance of approach to or deviation from the waypoint is simply visualized by drawing the waypoint on the navigation display. The pilot will realize functional behavior through turn maneuvers that turn the waypoint symbol right in front of the aircraft symbol. A locomotion model that enables heading changes is compatible with this visualization. For safety however, the simple presentation of intruder aircraft symbols on the navigation display only gives the pilot a mere notion of crashing and avoidance, not a meaningful perception. The visualization does not reveal which aircraft behavior avoids the intruder. Insight into how the motion of the own aircraft (locomotion) and the intruder realize the spatial separation, is obtained by considering the motions of the vehicles in a relative velocity plane that describes the own aircrafts motion relative to the considered intruder aircraft. The heading travel function, the locomotion model based on real turn dynamics (De Neef, 2002), calculates which turn maneuvers will cause a loss of separation in this plane and shows these turn maneuvers on the heading scale of the navigation display. The weakness of the guidance offered by the heading bands alone lies in the perception of efficiency goal related affordances. An off-line simulation proved that in a conflict situation, a resolution maneuver towards the closest heading band edge could lead to a larger lateral deviation from the original trajectory than a resolution maneuver to the other band edge that was situated further away. The perception of the angular proximity of the heading band edge and steering towards it, does not yield aircraft behavior that results in a minimal trajectory deviation.

In the relative speed plane, the relative velocity of the subject aircraft is described with respect to the considered intruder. A beam shaped area can be defined, outlined in Figure 3, by two lines originating from the own position and tangent to respectively the left and right side of the Protected Zone (PZ) of the intruder, at its present location. This zone is called the Forbidden Beam Zone (FBZ) and in Figure 3 the triangle indicates this zone. If the relative velocity vector is inside this area, the trajectory path will eventually enter the PZ and spatial separation will be lost.

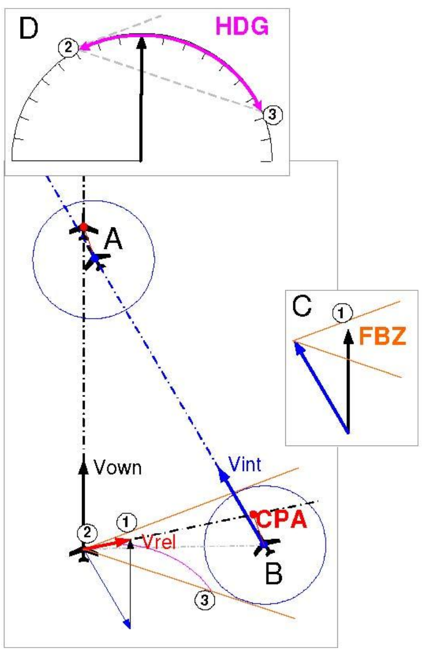


Figure 3. *Conflict Presentation in the absolute (A) and relative (B) velocity plane. Box (C) shows how the FBZ cuts out vector states. Index 1 shows a possible resolution state. Box (D) shows the heading band principle. Index 2 and 3 show the needed turn maneuvers. Vrel is the relative velocity with respect to the intruder aircraft. The circle around the intruder aircraft symbol is the Protected Zone (PZ)*

Separation can be realized by actions that will cause the relative velocity vector to lie outside the FBZ. As the relative vector is constructed by the own vector and the intruder vector, spatial separation can be realized by a vector state change (= aircraft maneuver) of the own vector, the intruder vector or a combination of those. Note that the magnitude of the relative speed vector is inversely proportional to the amount of time it takes until actual crashing or avoiding will take place. The origin point of the FBZ represents the point where relative velocity is zero. This means that when both aircraft have the same vector magnitude and heading, their relative position does not change in time. The Vector Envelope Map in Figure 3 shows all vector state possibilities that would assure separation.

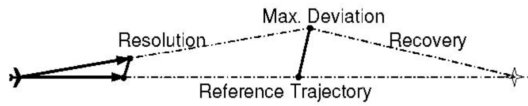


Figure 4. *Maximum deviation distance depends on the magnitude of the state change and the duration of the conflict resolution.*

The aircraft symbol at point "A" in Figure 3 shows the projected future location of the intruder aircraft at the closest point of approach. A visualization of this point does not lead to a useful display, since it will move considerably as avoidance maneuvers are performed; conflicts between aircraft have to be solved with heading and speed changes, and a presentation in absolute geometric space in this case does not provide the proper information to do this.

$$Deviation[m] = \left|V_{res} - V_{ref}\right| * t_{res} \quad (1)$$

Another issue to be considered is the efficiency of the chosen solution. Path deviation is quantified by the maximal spatial deviation (Figure 4). This is the distance between the actual and planned position of the own aircraft at the CPA instance. At that point the pilot will start the recovery maneuver in order to fly the aircraft back towards the original trajectory. The deviation due to conflict resolution is determined by two physical phenomena: the state change magnitude and the duration of the resolution or simply resolution time.

$$Deviation[m] = \left|V_{res} - V_{ref}\right| / \left|V_{res} - FBZ_{orig}\right| \quad (2)$$

Consider again Figure 3. The relative speed vector is constructed by taking the opposite of the intruder speed vector, and adding the speed vector of the own aircraft. In this graphical representation the end point of the relative speed is always lies at the end point of the absolute speed vector. Multiple conflicts can be combined in a single solution space by co-locating the end-points of the relative velocity vectors, as done in Figure5.

In the bottom part of Figure 5 one can see that the vector map presentation as it will be presented on the navigation display. The half-circles represent the maximum and minimum velocity boundary in which the pilot is allowed to operate. Also the heading change is limited to 90 degrees port and starboard in order to show travel opportunities that will yield destination approximation.

An aircraft maneuver of an intruder will be perceived by motion of the related FBZ. The pilot can directly act upon this motion if necessary. By steering in the opposite direction, a cooperative maneuver is realized with the perceived intruder maneuver. In a one-to-one conflict, the geometry of the envelope from the point of view of one aircraft is complementary to the other one. In Figure 6 one can see that moving against the direction of the other aircraft will cause the pilot to end up at the opposite FBZ leg. Furthermore, the closer one aircraft lies to one leg, the closer the other aircraft will lie to the opposite leg. Physically this means for example that, if one aircraft is close to the border that makes it pass the other at the left upper side, the other aircraft will be closer to the border line that will make it pass the first aircraft at the right lower side. In this way, cooperative maneuvers will always be initiated, even if both aircraft would begin their maneuver exactly at the same time.

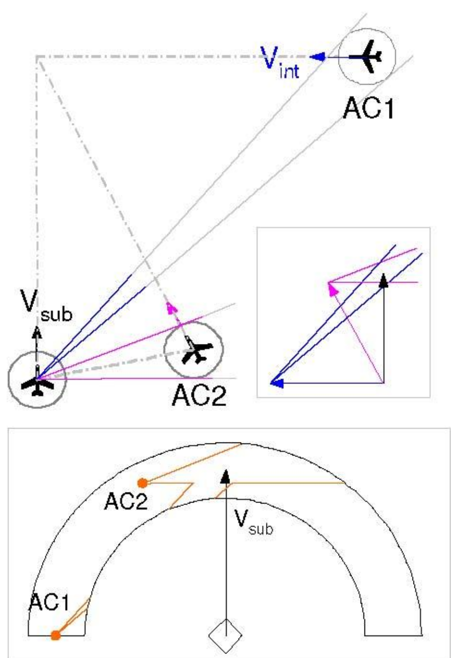


Figure 5. *Combination of the Forbidden Beam Zones for different conflicts, plotted in a vector map with the allowable heading and speed range, into a vector map showing heading and speed affordances.*

During the time that the own aircraft is approaching the intruder aircraft, the subject aircraft will get closer to the PZ and therefore the FBZ-beam will expand in time. The envelope presentation is based on direct state changes, so the geometrical form of the solution space does not take into account the beam expansion that evolves during the time period that the state change is realized. In Figure 7 a starboard turn maneuver is started by the subject aircraft at t(0) and is ended at time t(1). The solution state on the FBZ edge at the beginning of the maneuver will still lie inside the FBZ at time t(1),as the beam expanded during the time interval. The closer to the CPA instance, the more significant this phenomenon will become.

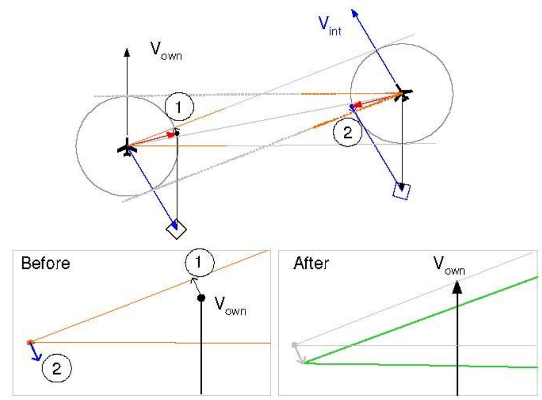


Figure 6. *Cooperative maneuvers of subject aircraft (1) and intruder aircraft (2).*

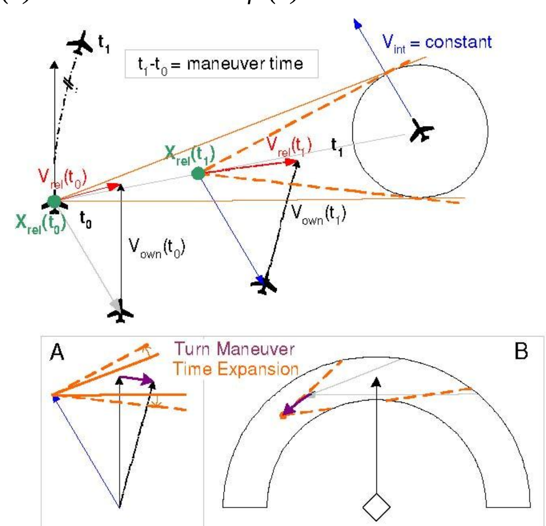


Figure 7. *Illustration of the time expansion of the FBZ during a resolution maneuver.*

Experiment

The state vector envelope principle was evaluated in a small a pilot experiment in a fixed-base flight simulator. Six pilots, aged 27 to 38, participated in the experiment, with experience ranged from a few hundred to three thousand flight hours. The pilots were asked to fly an IFR track between two waypoints in cruise flight conditions on an altitude of 30000 ft. At a given moment a conflict situation was detected and the pilot was asked to make a maneuver (using autopilot settings) that would result in a safe and efficient conflict resolution. When the intruder aircraft had passed by, the pilot began a recovery maneuver by going back to the original cruise speed and heading towards the next waypoint in order to continue its cruise flight on the original trajectory. They were instructed about the functioning of the speed vector envelope, and that the origin points of the other aircraft will yield a parallel trajectory at the same airspeed with the related intruder aircraft. During resolution the subject was allowed to change its strategy and to cross the forbidden zone to realize this change, as long as spatial separation with the intruder aircraft was maintained. Five different conflict geometries were simulated. No reference or other display designs were used, as the limited set up of this experiment investigates the feasibility of the newly designed guidance tool. First the pilots were briefed about the interface concept and the experiment design. Then 2 training runs and 5 experiment runs were done in the simulator. The training scenarios were similar but not equal to the experiment scenarios. Each run lasted 8 to 10 minutes. After the whole set of runs, the pilot was asked to fill in an evaluation form. The aircraft model used in the simulation was a Boeing 747-200. The aircraft was flying a cruise flight at 30000ft. Initial Velocity was chosen 0.8 Mach, about 240 m/s ground speed. The autopilot was enabled and IAS and heading could be manipulated on a virtual Mode Control Panel. The conflict algorithm for the subject aircraft, detects for the actual speed a future spatial separation violation of the 5 nm standard within 5 minutes look-ahead time. At his moment the envelope lines will be drawn on the display. Each intruder is simulated with a propagation model that defines an initial trajectory by its position, ground speed and heading. At a given time instance a resolution maneuver with a different ground speed and heading is triggered. When the intruders pass each other they will head back to their original trajectory path. The resolutions are human-like and will cause a spatial separation between 5 and 10 nm. The maneuver dynamics consists of simple turn geometry and a constant longitudinal acceleration. Both intruder aircraft only resolve the conflict situation with each other. In other words, they neglect the conflict situation with the subject aircraft. As a result it is possible that the intruder makes a counter-active or hostile maneuver. The occurrence of such an event makes it possible to check for robustness of the interface concept. In 26 out of 30 trials, the pilot's strategies were consistent with the rules for efficient solution of the conflict. In four trials an inefficient solution was chosen, solving the conflict but resulting in a large off-track distance. The behavior of the conflict aircraft was programmed with a pre-defined logic, however, resulting in a two runs with a loss of separation, due to “hostile” maneuvers of the conflict aircraft. Work is underway to provide the intruder aircraft with the proper behavior. All subjects indicated that the envelope interface was useful to them, but indicated that more training would be needed for an optimal comprehension of it. Their points of critique were on the actual implementation of the display (with lines instead of filled or shaded areas), and on the difficulty of correlating aircraft shown on the display with the shapes in the envelope. Another problem was to perceive the time left in a conflict. This was related to the FBZ expansion mentioned above, far from the conflict expansion is hardly noticeable, but closer by (when waiting too long with a solution) expansion would be rapid and prevent a reasonable solution. Two of the pilots quickly gained insight in the display, enabling them to predict and reason about the solutions well in advance.

Conclusions & Recommendations

The state vector envelope interface design is a guidance tool in the horizontal plane for the airborne planning of trajectory paths that maintain spatial separation with other aircraft, approach the destination and limit path deviation while resolving a conflict situation. A locomotion model based on instantaneous combined speed and heading changes describes aircraft motion in a way that it matches flight practice. The realization of the trajectory planning task is based on the pilot perception of travel-relevant airspace affordances like crashing, avoiding, approaching and deviation in terms of combined heading and speed changes. The state vector envelope presentation especially visualizes the affordance of collision & avoidance by the envelope lines and the affordance of path deviation minimization by the envelope origin points. A simple and effective rule and skill based conflict resolution strategy consists of steering out of the forbidden vector zone while avoiding the state vectors of other aircraft. A simple experiment with two intruder aircraft showed that in most occasions the pilot conflict resolution behavior matches with the expected resolution strategy. The pilot feedback underlined that the envelope concept is useful, but more study and training is needed to get more insight in the conflict geometry presented. Furthermore it was difficult to perceive intruder maneuvers and to correlate an intruder aircraft with its respective part in the envelope form. The most important shortcomings however, are the lack of urgency awareness and the expansion of the beam width. It is difficult to predict when the subject aircraft will pass or crash into an intruder aircraft. Combined with the expansion phenomenon, this means that the pilot does not know how much time is left to resolve a conflict, neither how much the envelope edges will expand during the resolution maneuver. The use of different "urgency layers" for the envelope form and the presentation of the "time to impact/avoid" give a notion for urgency. The beam expansion could be faced by plotting a future prediction of the envelope form. The best remedy however, is to upgrade the locomotion model from instantaneous state changes to realistic maneuver dynamics. Currently, work is underway to improve the presentation of the vector envelopes, and perform a more elaborated evaluation. Future directions could be the extension to 3D navigation, i.e., including altitude; however, this poses some challenges regarding the visualization of the affordances. Further improvements could be inclusion of the turn and acceleration dynamics, as this would address the uncertainty about beam expansion, and it would also make the interface more generally applicable to other vehicles.

References

ACM Airborne conflict management: Application description v2.5 (2002), RTCA, Federal Aviation Authorities, Tech. Rep. RTCA SC-186.

Hoekstra (2001), J., *Designing for safety: the free flight air traffic management concept*, Ph.D. dissertation, Delft University of Technology. Amsterdam: NLR. ISBN 90-806343-2-8.

van Paassen, M.M., (1999), *Functions of space and travellers*, in XX European Annual Conference on Human Decision Making and Manual Control, Loughborough, United Kingdom.

De Neef, R., (2002), *Functional modeling of airspace: Definition and evaluation of a travel function for application to aircraft heading guidance*, Unpublished Msc. thesis, Faculty of Aerospace Engineering, Delft University of Technology.

Modarres, M., *Functional modeling applications association website*. http://www.enre.umd.edu/ifmaa/.

Gibson, J.J. (1979), *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin.

Vicente, K.J. and Rasmussen, J. (1992). *Ecological Interface Design: Theoretical Foundations*. IEEE Transactions on Systems, Man and Cybernetics, Vol. 22, No. 4, pp. 589-606

Lind, M. (1990), *Representing goals and functions of flow modelling - an introduction to multilevel flow modelling.* Institute of Automatic Control Systems, University of Denmark, Tech. Rep. 90-D-381.

“MEANINGFUL PHYSICS”
OR FINDING A SYSTEM DESCRIPTION SUITABLE FOR ECOLOGICAL INTERFACE DESIGN

M. M. (René) van Paassen
Max Mulder
Stijn B. J. van Dam
Aerospace Engineering, Delft University of Technology Delft, The Netherlands
Matthijs Amelink
DECIS Lab
Delft, The Netherlands

Ecological Interface Design (EID) is a design paradigm that addresses the cognitive interaction between users and systems. EID’s original application domain is the field of process technology. However, in several studies the techniques outlined for EID are applied to other domains. In the development of EID interfaces for two different tasks involving control of the locomotion of an aircraft, the authors experienced a gap between the stages of cognitive work analysis (CWA) and the actual design of the interfaces. This paper analyses the approach in the two projects, generalizing the findings in creation of a proper representation with the Abstraction Hierarchy (AH) identified in the CWA stage. For these, and probably other projects, it appears advantageous to consider alternative and possibly parallel expressions for the constraints identified at the Abstract Function level of the AH, to create a match between either user controls and the representation, and between system purpose and the representation.

Introduction

Ecological Interface Design (EID) is a design paradigm that addresses the cognitive interaction between users and systems. EID’s original application domain is the field of process technology. However, in several studies the techniques outlined for EID are applied to other domains.We have recently developed EID interfaces for two different tasks involving control of the locomotion of an aircraft (Van Dam, Abeloos, Mulder and van Paassen, 2004; Amelink, Mulder, van Paassen and Flach, 2005). In these projects, we experienced a gap between the stages of cognitive work analysis (CWA) and the actual design of the interfaces, which in both projects was bridged in a serendipitous manner. In both cases, the results from the CWA alone were not a sufficient starting point for the development of the EID interfaces, an additional system description was needed.

In the first project, (Van Dam et al., 2004), the task of self-separation in a free-flight environment was considered. At an abstract function level, maintaining a sufficient distance from surrounding vehicles was identified as the basic principle by which safety was achieved. Computer algorithms exist, and are being used, to determine whether for the current flight trajectory sufficient distance from other vehicles is maintained. Such computer-based methods serve to identify future separation problems, and can be used to explore the solution space available. However, to the user, interfaces based on such algorithms appear to present “black box” solutions, and it is not immediately clear how a certain control action leads to achievement of sufficient separation. A meaningful representation of the problem, and thereby the EID design, was only possible after realizing that separation between two moving vehicles is achieved or destroyed by relative movement. This led us to explore the aircraft motions in the relative space, rather than in geodetic coordinates. The exploration proved to be the key that led to the interface design.

The second project, considered the task of following an altitude profile with an aircraft (Amelink et al., 2005). In contrast to the other problem, here the evaluation of the aircraft’s dynamics against the set criteria is clear, but the relation between the control actions and the aircraft’s response is less obvious. Again the solution was exploration of the aircraft motions in a different representation space, in this case in terms of kinetic and potential energy. The motivation was different, here the simpler relationships between control inputs of the system and the outputs in terms of energy motivated the choice of a different system representation. The criteria, the height and velocity profile, could be re-formulated in terms of energy and presented in the display.

The essential step is finding alternative system descriptions that match either the criteria and solution space, or more closely match the available controls (affordances). We termed this representation “meaningful physics”, since the representation must not only be physically correct, but also compatible with the human's goal oriented behavior. This paper shortly discusses both projects, and elaborates on the common elements in the approaches.

Avoiding traffic

The growing intensity of air traffic leads to high workloads and congestion, not only at airports where aircraft need to land and take off again, but also in the air traffic system. The main task of an air traffic controller in en-route airspace is the separation of aircraft, and, depending on the structure of the airspace, there is a limit on the number of aircraft that a controller can handle. Reducing the size of sectors is not a valid solution, since this increases the coordination required for passing aircraft from one sector to another. Various studies indicate that, with the proper support, the separation task in en-route airspace can be delegated to the flight deck. This would allow *direct routing*, in which aircraft fly a trajectory straight to their destination instead of via designated airways, and *cruise climb*, in which the aircraft flies at the most economical altitude at all times.

Present systems, such as the ASAS (Airborne Separation Assurance System) and pASAS (predictive ASAS) systems developed by the Netherlands Aerospace Laboratory NLR (Hoekstra, 2000), have proved to offer the pilot a safe and effective conflict detection and resolution. However, these systems have an advisory nature; a computer algorithm determines the possible solutions, and presents these to the pilot, whereupon the pilot can choose to implement one of the solutions. In our design, we intended to make an interface based on EID principles, which would show the *situation* to the pilot in such a manner that the solutions to a conflict would appear obvious and logical.

Workspace analysis As for most systems, three goals for a traveling vehicle can be identified at the functional purpose level, production, economy and safety (Figure1). When considering the locomotive aspects of the problem alone, i.e. ignoring issues such as staying within the flight envelope, assuring propulsion, lift, atmospheric protection etc., the primary principle for achieving safety is maintaining separation from potentially hazardous objects, such as other vehicles and stationary objects. For an aircraft this means that other aircraft and terrain must be avoided during flight, or, in other words, it needs to maintain separation. Separation can be predicted when an estimate of the trajectory of the own aircraft and of other aircraft in the vicinity is known, but this prediction is not equal to a mapping of the affordance, as experiments with early systems such as ASAS (Hoekstra, 2000) have shown. These systems show the conflict, focusing on the location in the surrounding space where the conflict will be, however, they do not show how to avoid the conflict, since, as one tries to maneuver away from the predicted conflict location, the conflict location will changes, and also new conflicts may be created. The main challenge is expressing the (expected) motion of the own craft and other aircraft in the vicinity in such a manner that the affordances (of crashing or avoiding) are clearly visible.

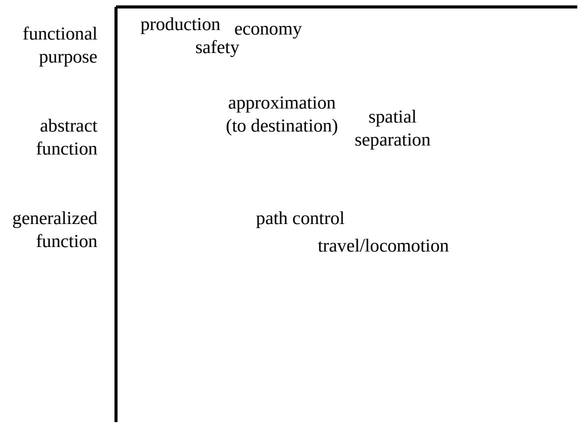


Figure 1: Abstraction hierarchy for flight, only for the aspects of traveling toward the destination and conflict avoidance.

At the abstract function level, the system was described in terms of the kinematics of travel. As in the well-known prototype system DURESS(Vicente and Rasmussen, 1990), this level reflects the laws of physics acting on the system; thermodynamics and mass balances in the case of DURESS, versus kinematics and locomotion in the present case. The dynamics and limitations of the turn needed to avoid other traffic was neglected in this analysis, and thus the kinematic equations for travel over the earth's surface can be given as:

$$\dot{F}_T = \frac{V_{north}}{R_M + h} \quad (1)$$

$$\dot{l} \cos F_T = \frac{V_{east}}{R_P + h} \quad (2)$$

With F_T and l_t as latitude and longitude, R_M and R_P as the radii of curvature fitting the earth ellipsoid to a meridian section and east-west section respectively, h

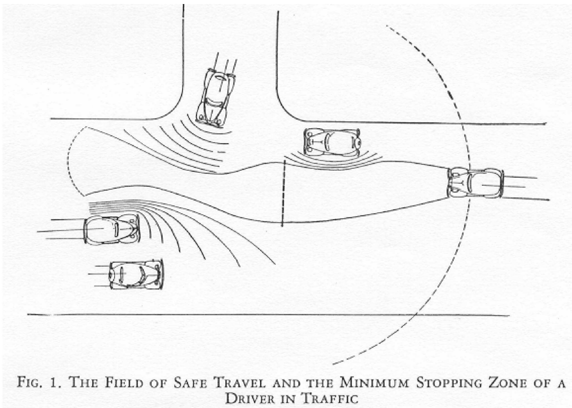


Figure 2: Visualization of the field of safe travel, from (Gibson and Crooks, 1938)

as the altitude. This is a perfectly adequate description enabling travel in a certain direction. It can be visualized in a moving map display, as for example is done for most electronic navigation displays in current aircraft. One can say that the representation is correct and complete within the requirements of accuracy. However, it does not show the path to avoid obstacles (aircraft) *moving* in that same environment.

An example of such possible paths is given in the illustrations in Gibson's 1938 paper(Gibson and Crooks, 1938). The intuitive looks of the solution are deceiving, since at each instant there is an infinity of directions and speeds to choose from, and paths much more bizarre than sketched in the picture are certainly possible. The same is possible for the aircraft avoidance problem. Presenting all possible control actions and future paths is thus not feasible. In order to keep the solution space acceptable to human pilots, only maneuvers that consist of a single turn to a new heading, possibly combined with an increase or decrease in speed, are considered.

A simple and enlightening presentation can however not be created from the navigation equations (2). A different way of expressing aircraft motion was needed, in this case by using the "intruder" aircraft as an origin, and expressing all motion relative to that (moving) reference frame. Kinematic constraints, such as the turn dynamics, need to be translated to this reference frame (De Neef and van Paassen, 2001). Here first the case is considered where these dynamics can be neglected, ongoing research focuses on the inclusion of some of these dynamics. In a reference frame with the intruder at the origin, the speed of the own aircraft is the relative speed with respect to the intruder,

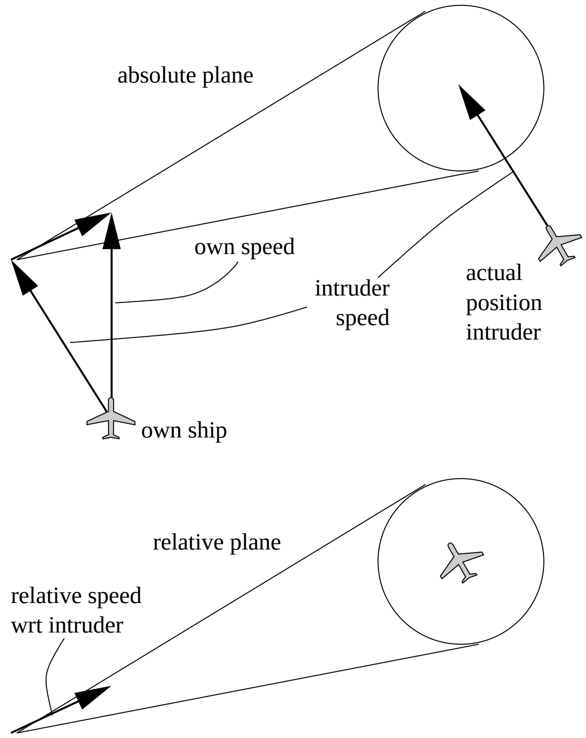


Figure 3: Overlay of the relative and absolute movement spaces.

and avoidance or intrusion is simple to check (Figure 3). Moreover, by overlaying the relative reference frame onto the absolute reference frame, the relation between the aircraft's absolute velocity and the relative velocity with respect to the intruder is shown. A picture of the display as tested in a simulation (Van Dam et al., 2004) is shown in Figure 4.

Keeping path and speed

The second project discussed here grew out of a curiosity about the strategies pilots would use for correction of deviations in speed and altitude in an approach to the runway. Initially, two perceived strategies were considered:

- "Throttle to speed, stick to altitude". In this strategy, which is also implemented in most current autopilots, elevator control inputs are used to correct altitude deviations, and the throttle is used to correct speed deviations.
- "Stick to speed, throttle to altitude". The reverse of the above strategy. Control theoretic analysis shows that this strategy would also work for an

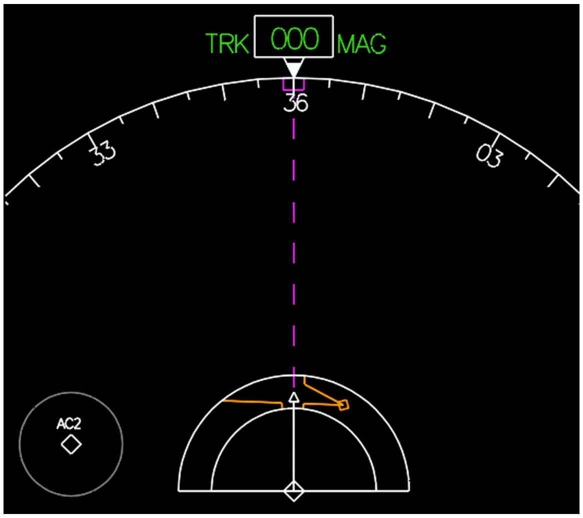


Figure 4: Display for avoiding aircraft traffic. A speed and heading space is added to an electronic navigation display. The two white half-circles show minimum and maximum speed, the orange shape shows speed and heading combinations that lead to conflicts.

autopilot, albeit with less performance and stability.

Also in this project an approach with cognitive work analysis was followed. The work domain in this case had a limited scope, with an altitude reference path and a speed reference at the functional purpose level.

Workspace analysis In the development of the abstraction hierarchy and in an engineering analysis of the problem, the energy balance of the aircraft proved to be a crucial element (Figure 5). The principle of an energy balance is also used in "new" designs for autopilots that function on the principle of energy controls (Lambregts, 1983a; Lambregts, 1983b). However, practice has shown that conventional autopilots, using controllers based on feedback of altitude to the elevator and feedback of speed via the throttle, also function in a satisfactory manner. This kind of feedback would be based on a Newtonian view of the measured kinematics, where the airplane is considered to have inertia that can be accelerated along its path to achieve speed control or perpendicularly to its path to achieve altitude control.

An alternative formulation of the aicraft kinematics, in terms of kinetic and potential energy, is also an option. The advantage of the energy formulation is that it more closely corresponds to the effect of the control inputs, with the throttle influencing the total energy sum, and

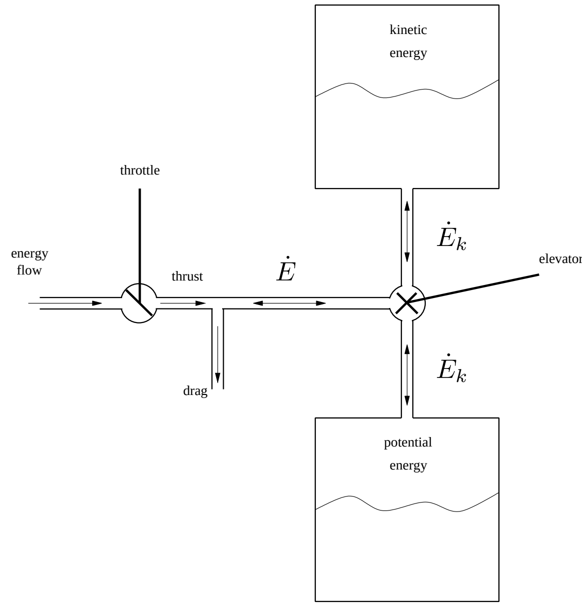


Figure 5: Aircraft altitude and speed control can be visualized as an energy balance, where throttle and drag determine the change in total energy level and the elevator determines the distribution between kinetic and potential energy.

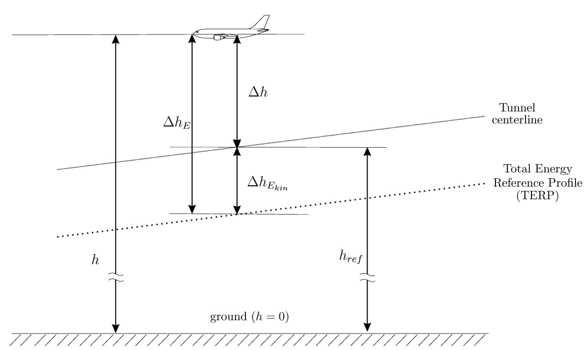


Figure 6: Instead of a target speed and altitude, the functional in the altitude and speed control task are expressed as a total energy reference plane (TERP), and a potential energy reference plane

the elevator, via control of the flight path angle, balancing the rates of potential and kinetic energy.

The automatic control system that is based on the energy balances; the “total energy controller”, is claimed to have a better performance and to need less tuning than the conventional controllers. However, the signals that are presented to the pilot in current flight deck designs correspond to the feedback signals used by conventional autopilots. In a design for the energy display, the target values that were compatible with the energy formulation were added to a perspective flight path display(Amelink et al., 2005), allowing human pilots to also (or more easily) adopt the control strategy of the total energy controller. An illustration of the total energy plane in relation to the altitude (kinetic energy) plane is given in Figure 6.

Reflection and Conclusion

Reflecting on both projects, and on the approaches taken in other projects, a number of common elements in both approaches are found. The first element is that a certain fast part of the dynamics of the controlled system does not need to be represented in an ecological interface.

A certain portion of the system dynamics can be too fast, or too trivial, for presentation on a display. In the case of aircraft altitude and speed control, the attitude dynamics of the aircraft are not presented in the display. They need not be, since a pilot’s basic training enables her/him to handle these dynamics. So, instead of using the true control input (yoke position) to the aircraft, the aircraft attitude and flight path are the control input in the portion of aircraft dynamics and kinematics considered, relying on the pilot to implement these.

A second element is that, in order to make the task acceptable to the operator, it may be necessary to reduce the potentially high dimensionality of the solution or input space. For the aircraft avoidance problem, this resulted in choosing a single maneuver to a new heading. Again, the capacity to turn the aircraft to a selected heading was trusted to the pilot in this case.

And finally, it is often necessary to consider different formulations for the “physics” at the abstract function level. Either because the additional formulation provides a better match with the controls, as in the second example, or because it provides a better way for expressing the achievement of the functional purpose, as in the first example.

References

Amelink, M. H. J., Mulder, M., van Paassen, M. M. and Flach, J. M. (2005). Theoretical foundations for a total-energy based perspective flight-path display, *The International Journal of Aviation Psychology* . Accepted for publication.

De Neef, R. and van Paassen, M. M. (2001). Functional modelling of airspace, *in* M. Lind (ed.), *XX European Annual Conference on Human Decision Making and Manual Control*, Technical University of Denmark, Kgs. Lyngby, Denmark, pp. 71–80.

Gibson, J. J. and Crooks, L. E. (1938). A theoretical field analysis of automobile driving, *American Journal of Psychology* **51**(4): 453–471.

Hoekstra, J. e. a. (2000). Overview of nlr free flight project 1997-1999, *Contract Report NLR-TP-2000-227*, National Aerospace Laboratory NLR.

Lambregts, A. A. (1983a). Functional integration of vertical flight path and speed control using energy principles, *1st Annual NASA Aircraft Control Workshop NASA LRC*.

Lambregts, A. A. (1983b). Operational aspects of the integrated vertical flight path and speed control system, *SAE Technical Paper Series* **831420**.

Van Dam, S. B. J., Abeloos, A. L. M., Mulder, M. and van Paassen, M. M. (2004). Functional presentation of travel ipportunities in flexible use airspace: An EID of an airborne conflict support tool, *IEEE International Conference on Systems, Man and Cybernetics*, The Hague, pp. 802–808.

Vicente, K. J. and Rasmussen, J. (1990). The ecology of human-machine systems ii: Mediating “direct perception” in complex work domains, *Ecological Psychology* **2**(10): 207–249.

FACIAL TEMPERATURE AS A MEASURE OF MENTAL WORKLOAD

Hans (J.A.) Veltman
Wouter (W.K.) Vos
TNO Defense, Security and Safety
Soesterberg, The Netherlands

We conducted an experiment to explore the relation between facial temperature and mental effort. Participants had to perform mentally demanding tasks while their face was captured with an infrared camera. The temperature of the nose decreased during these tasks and increased during the successive rest periods. Other parts of the face did not change due to mental effort. The advantage of this workload measure is that it can provide objective and real-time information about mental effort of operators without attaching sensors to an operator. This measure can be used to measure the workload of operators in relatively stable environmental conditions such as air traffic controllers and operators of unmanned aerial vehicles.

Introduction

The task of air traffic controllers and operators of unmanned aerial vehicles can be highly demanding from time to time. Although people can adapt efficiently to changing task demands, a high mental workload does have negative effects such as an increased likelihood of human error.

There are many different techniques to measure mental workload. These techniques can be categorized into performance measures, subjective measures and physiological measures. These types of measures do not provide the same information (Veltman & Jansen, 2003). Performance is often difficult to measure in applied situations and when it can be measured, it often does not provide adequate information. Operators have the ability to adapt to changing task demands by investing more effort and therefore, an adequate level of performance can often be maintained at the cost of high workload. If the workload becomes too high, the performance often decreases dramatically. It is important to have information about the state of an operator before the level of performance decreases. Subjective workload measures provide more information about the workload but these measures are also difficult to obtain in applied situations. Finally, physiological measures mainly reflect the amount of mental effort that an operator has to invest in order to perform the task adequately. They can provide continuous and objective information about the state of an operator. This is necessary if one wants to prevent a decrease of performance (Hockey, 2003).

An important disadvantage of physiological measures is that most often electrodes or other sensors have to be attached to the person, which restricts the use of these measures in many applied settings. The measurement of facial skin temperature by means of an infrared camera might not have this practical limitation. There are some indications that the face temperature, especially the temperature at the nose, decreases when mental workload increases (e.g. Genno et al., 1997).

In this paper we describe an experiment in which the applicability of facial temperature for the assessment of mental workload is further explored. In this experiment participants had to perform mentally demanding tasks during which the facial temperature was measured with an infrared camera. We explored if the facial temperature changes due to mental effort. Moreover, we explored the most sensitive locations on the face and the sensitivity to different levels of task load.

This experiment is part of a research program in which the possibility for adaptive automation is investigated. Adaptive automation is a concept in which the level of automation is adapted to a specific situation. The state of the operator can provide relevant information for adaptive automation such as a high workload of the operator. If the mental workload of an operator is too high, the overall performance might increase if the taskload is reduced. This can be accomplished for example by taking over some tasks from the operator, present some tasks to another operator, or wait to present less relevant information until the workload of the operator is reduced.

Facial temperature seems to be a promising element in adaptive automation concept because it might provides objective information about the workload of an operator and it can be obtained relatively easy.

Method

Participants

The experiment has been performed on eight participants, six males and two females, their ages ranging from 23 to 41.

Task

The participants had to perform three mentally demanding tasks. We used the auditive version of the Continuous Memory Task (CMT) that has been shown to be a highly cognitive demanding task in earlier experiments (e.g. Veltman & Gaillard, 1998). This task is mentally demanding because the participants have to compare each letter with the letters from the memory set and more important, they have to use their working memory continuously.

The participants had to remember two or four target letters (CMT2: A-B and CMT4: A-B-X-Y). Letters from the alphabet were presented randomly with an interval time of three seconds. About 30% of the letters were targets. The participants had to press a button when they heard a target letter and press another button when they heard the letter for the second time. Thus, they had to react to the target letters and had to count them independently.

The word “okay” was presented after a correct response and the word “nope” after an incorrect response and after an omission. The participant had to restart counting after feedback was provided.

The duration of each task was three minutes. Before and after each task there was a rest block of three minutes (see Figure 1).

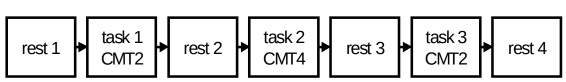


Figure 1. *Scheme of the experimental conditions. Each block lasted three minutes. CMT2 is the Continuous Memory Task with two target letters and CMT4 is the Continuous Memory Task with four target letters.*

Procedure and Apparatus

Before the experiment the participant was trained in the CMT: one minute training for the two-target letter task and one minute of training for the four-target letter task. The training was primarily meant to inform the participant about the goal of the task. The participants were told that they had to make fast responses and had to avoid errors.

The task was meant to induce a high mental effort only. The performance on the task was not relevant and therefore, this will not be presented.

We used a FLIR SC 2000 infrared camera that was able to take temperature pictures with an accuracy of 0.07 °C (14 bit). The resolution of the camera was 320 x 240 pixels. The camera took pictures with an interval time of 5 sec. So there were 36 pictures of the face in each block.

Since it was expected that the nose temperature would be the most interesting area, four points on the nose were manually selected in the first picture of each participant in Matlab 6.5.1. Another 13 points were selected on the rest of the face (see Table 1).

Table 1. *Measurement locations on the face.*

NR.	Location
1	Middle forehead
2	Left side forehead
3	Right side forehead
4	Upper inside left eye
5	Lower inside left eye
6	Outside left eye
7	Upper inside right eye
8	Lower inside right eye
9	Outside right eye
10	Nose bridge
11	Left side nose
12	Right side nose
13	Nose tip
14	Left cheekbone
15	Right cheekbone
16	Between nose and upper lip
17	Between lower lip and chin

A chin rest was used to stabilize the head. Despite this chin rest, the participants were not able to keep their head at the same position throughout the experiment. Therefore, a procedure was developed to match the selected points in each picture. The seventeen points were selected in the first picture with mouse clicks. A rectangle around the nose was also selected in this picture. This rectangle was correlated in the X and Y-axis with a larger rectangle in the consecutive pictures (20 pixels larger at each side of the original rectangle). Based on this two-dimensional correlation, the points in consecutive pictures were shifted. The adjusted positions were visually checked and it appeared that all selected points remained to their place relative to the head. The correlation procedure helped us much in the present experiment. However, it will only work for minor changes of the head position. In a situation in which operators can move their head freely, a more elaborate procedure to get fixed locations of the face has to be developed.

Data Analysis

In each block we had 36 temperature values for each location. The temperature of each location consisted of an average in a circular area with a diameter of 5 pixels (≈ 7 mm). Within each block, we fitted straight lines (least squares method) through the data points for each location. Measurements outside ± 3 times the standard deviation range were removed. This resulted in 17 (locations) x 7 (blocks) regression lines for each participant.

We used an ANOVA repeated measurement analysis to analyze whether differences in temperature were significant. The following comparisons were made:

- differences between the seven experimental blocks (one factor with seven levels);
- differences between the three task blocks (one factor with three levels);
- differences between the four rest blocks (one factor with four levels);
- differences between the three locations at the nose that appeared relevant in the previous analysis. For this analysis we calculated the average value of the three rest blocks and the average values of the three task blocks and tested whether the temperature changes were different at the three locations.

Differences were further explored with post hoc analysis (Tukey HSD).

Results

Figure 2 presents an example of the data for one participant. The temperature at the forehead and the left side of the nose is plotted for each camera frame. The regression lines are also plotted in this figure. The slopes of these regression lines were used for statistical analysis. This picture shows that the temperature at the nose decreased during tasks and increased during the following rest period. The temperature at the forehead is almost stable for this participant. The temperature of the nose changed substantially during a rest and a task block.

The change of the nose temperature started almost directly after the start of a block. This indicates that it is a fast reacting measure. The data of the other seven participants showed similar patterns, but the average range was lower than the data in Figure 2.

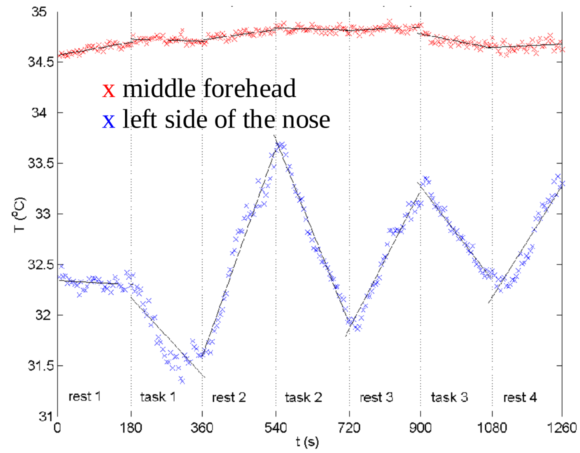


Figure 2. *Measurements of the temperature around two locations on the face (forehead and nose) of one participant. For each period straight lines were fitted through the data with the least squares method.*

Figure 3 presents the average temperature slopes for all locations. The rest blocks are presented separately from the task blocks in this figure. The strongest differences between the blocks were found at the left side of the nose [(F(6,42)=12.25, p<0.001)], at the right side of the nose [(F(6,42)=10.81, p<0.001)] and at the nose tip [F(6,42)=9.57, p<0.001]. During the tasks, the temperature decreased substantially for these locations and during the rests the temperature increased. Post hoc analysis revealed that the temperature slopes during all three tasks differed significantly from all four rests for the three locations at the nose (location 11, 12 and 13).

The temperature decreases during the three tasks was different for the left side of the nose [F(4,14)=6.6, p<0.01] and for the nose tip [F(4,14)=4.86, p<0.05]. Post analysis revealed that the temperature decreased more during CMT4 (second task) than during the first CMT2 task.

Figure 3 also shows a rather small decrease in temperature at the upper lip (location 16) during task2 and task3 and an increase during rest2, rest3 and rest4. Statistical analysis revealed that only the temperature during task2 was different from rest2. No other statistical effects were found for this location.

Some other significant effects were found, but these effects were very small and were not systematic. Therefore, they are not described here. The smallest temperature changes were found at the forehead. These differences were far from significant.

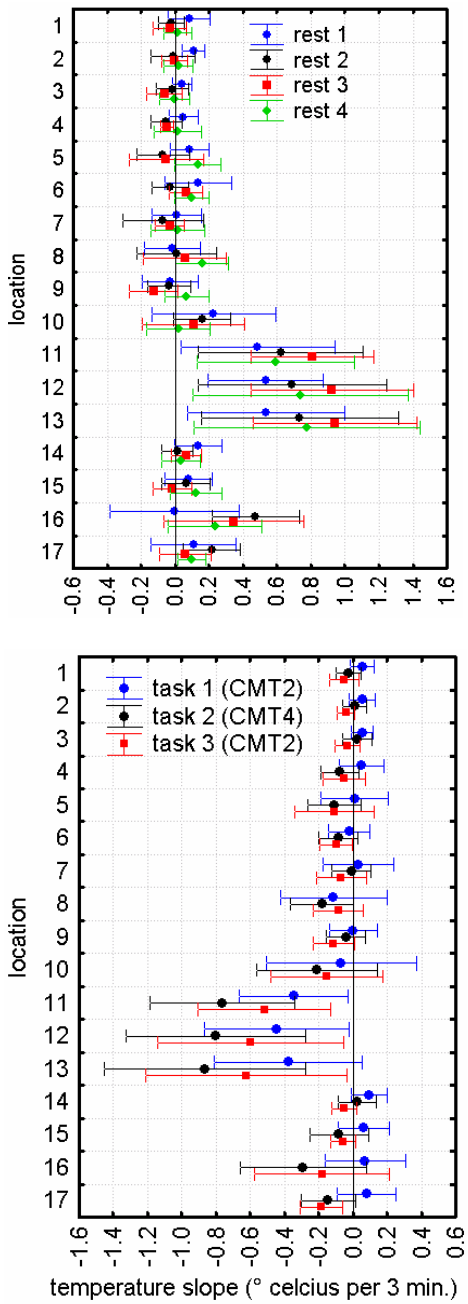


Figure 3. *Average temperature change for each measurement location. The top figure shows the temperature changes during rests and the bottom figure shows the temperature changes during tasks.*

Discussion

The results of this experiment show that mental effort can be distracted from changes in temperature of the nose. There was a clear difference between the temperature change during tasks and during rests. We found a small difference between the nose temperature during the four-target CMT (task2) and the temperature during the two-target CMTs (task1 and task3). The temperature decrease during the four-letter CMT was stronger than during the first two-letter CMT for the left side of the nose and the nose tip. A smaller and not significant difference was found between the four-letter CMT and the second two-letter CMT. This indicates that there is a relation, albeit not a strong one, between the amount of mental effort investment and the decrease in nose temperature. It should be noted that a higher task demand does not necessarily result in an increased mental effort. It is possible that some participants already did their utmost best during the two-target CMT task. The more demanding four-target CMT task would not have increased their effort investment anymore.

The decrease in nose temperature is most probably due to a dilation of the veins in the nose. This causes a reduction of blood flow and as a consequence the nose temperature will adapt faster to the environmental temperature. The diameter of the veins in the nose is mainly regulated by the sympathetic part of the autonomic nervous system (Widdicombe, 1993; Lung, 1995). Mental effort causes a reduced para-sympathetic activity and an increased sympathetic activity. This is the reason why most physiological measures, such as cardiovascular measures are sensitive to mental effort.

An increased sympathetic activity is probably not the only cause of the decrease in temperature. Mental effort often results in increased ventilation (Wientjes, 1993; Veltman et al., 1998). More relative cold air might flow through the nose during tasks compared to rests and therefore, the nose temperature drops. The data of the present experiment indicate that respiration might be involved in the present results. The heads of the participants were fixed with a chin rest, which forces them to breathe through the nose. This causes a flow of air around the nose. We found a small difference in temperature on the measurement location between the lip and the nose. Although these differences were not statistically significant, it indicates that respiration does play a role. Further experiments must clarify what the exact mechanisms of the temperature change are. However, for the application of this measure it is less relevant, because both mechanisms result in a temperature decrease during mental tasks.

There are several applications for workload measures. One of the applications is the evaluation of interfaces with regard to differences in mental effort investment. The nose temperature can be used for this, because there are several participants involved in such tests. If a few participants do not show changes in nose temperature, this does not affect the outcome very much.

Another application is adaptive automation for which it is very important to have highly reliable information about the effort investment. Operators will never accept that a system will take over tasks or delay less relevant tasks based on incorrect measurements. The reliability of nose temperature is too low to be used in adaptive automation. The reliability of the effort measured can be increased if the nose temperature is combined with other measures. Preferably this should be measures that do not require sensors to be attached to the operator or measures that are obtained with wireless sensors that do not hinder the operator in performing tasks. Examples of such measures are wireless heart rate sensors and eye point of gaze measures.

References

Genno, H., Ishikawa, K., Kanbara, O., Kikumoto, M., Fujiwara, Y., Suzuki, R. et al. (1997). Using facial skin temperature to objectively evaluate sensations. *International Journal of Industrial Ergonomics., 19,* 161-171.

Hockey, G. R. J. (2003). Operator Functional State as a Framework for the Assessment of Performance Degradation. In G.R.J. Hockey, A. W. K. Gaillard, & A. Burov (Eds.), *Operator Functional State: The assessment and Prediction of Human Performance Degradation in Complex Tasks.* (pp. 8-23). Amsterdam: IOS Press, NATO Science series.

Lung, M. A. (1995). The role of the autonomic nerves in the control of nasal circulation. *Biological signals, 4,* 179-185.

Veltman, J. A. & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics, 5,* 656-669.

Veltman, J. A. & Jansen, C. (2003). Differentiation of mental effort measures: consequences for adaptive automation. In G.R.J. Hockey, A. W. K. Gaillard, & O. Burov (Eds.), *Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks.* (pp. 249-259). Amsterdam: IOS Press.

Widdicombe, J. (1993). The airway vasculature. *Experimental Physiology, 78,* 433-452.

Wientjes, C. J. E. (1993). *Psychological influence upon breathing: Situational and dispositional aspects.* Thesis, University of Brabant.

ANALYSIS OF ADVANCED AIRSPACE CONCEPT OPERATIONS USING HUMAN PERFORMANCE MODELING

Savita A. Verma
NASA-Ames/SJSU
Moffett Field, California

Kevin M. Corker
San Jose State University
San Jose, California

The Advanced Airspace Concept (Erzberger, 2001) proposes to achieve increased capacity in both en route and terminal areas through the use of technologies that include air-ground datalink, automation generating 4-D trajectories, and an independent back-up system, intended to provide safe transition whenever there is a malfunction. An analysis of the concept's operations was performed using the human performance model Air Man Machine Integrated Design Analysis System (Air MIDAS) (Corker, 2000). For this research, three types of operations were modeled for an Air Traffic Control (ATC) agent – current operations, Automated Airspace Concept (AAC) operations, and Tactical Separation Assisted Flight Environment (T-SAFE) operations.. The results suggest that AAC operations decreased controller's workload when compared with current day operations. However, transition of the aircraft from AAC through T-SAFE to standard ATM control increased workload for the period of transition. This was marked with a high level of activity for the ATC-agent under the current and T-SAFE operations as the ATC agent sought to update its internal world representation with relevant aircraft trajectories to assume manual control.

Introduction

Various research efforts are focused on increasing the capacity of National Airspace (NAS). Advanced Airspace Concept (AAC) proposed by Erzberger (2001) is one such future concept, whose objective is to reduce Air Traffic Controller's (ATC) workload by automating the tactical functions of the ATC and by providing near term separation assurance. The key to safety is defined in the concept through the redundant system called Tactical Separation Assisted Flight Environment (T-SAFE).

The core ideas of the concept include segregation of the airspace into two categories, one is advanced operations airspace, and the other is standard or current practice airspace. The advanced airspace/sectors will combine several current sectors into a large single sector airspace, and this airspace configuration will be used during peak periods. Handoff between advanced airspace and standard airspace will be done using automation. Also, there are two kinds of aircraft anticipated in the concept, equipped and unequipped. The equipped aircraft are distinguished by their ability to exchange 4-D trajectories with the ground system and follow them accurately.

The ground automation called the Advanced Airspace Computer System (AACS) will generate 4D trajectories that will provide all the equipped aircraft with conflict free trajectories via data link. Flight crew reviews these trajectories before they are downloaded into the FMS. In the case that the aircraft does not accept a clearance issued by the ground system without cause, then that aircraft will be handed off to the manual system, i.e. to the ATC.

The concept has also has a fail-safe system , T-SAFE, which is an independent back up system that runs in parallel with the AACS, and is intended to provide a safe transition between automated and manual operations, in the case that AACS fails. Thus T-SAFE independently verifies that every trajectory provided by AACS is conflict free for the next 3 minutes before uplinking to the aircraft.

The present research effort hypothesized that workload levels for the ATC agent would be the highest in the T-Safe mode. This was based on previous research done on workload and mixed equipage. Several other studies (Corker, Gore, Flemming and Lane, 2001 or Jara & Corker, 2002) on mixed equipage operations have found that mixed operations can be challenging to the controllers. They found that highest subjective workload was reported by the operators in the mixed equipage mode. For example ATC reported higher subjective workload in the 80% free flight versus 100% free flight in the study by Corker, Gore, Flemming & Lane (2001). The human performance model used to analyze the impacts of the concept on human performance, particularly workload, is described in the next section.

Human Performance Model

Air Man Machine Integrated Design and Analysis System (Air MIDAS) is a computational human performance model used to predict human performance in joint cognitive systems. The system has been used in various environments ranging from aviation and emergency response systems to military systems (e.g., Corker & Smith, 1992). It has agent-based architecture and represents the physical world (e.g., equipment and terrain) and human perception – attention, and other cognitive functions – to varying degrees of fidelity. The system can represent a large number of human agents. Each human agent has, at its core, an internal representation of the physical world, a scheduler, and task demands. The dynamic interplay of all these components represents human interaction with automation. Some of the components or constructs represented in Air MIDAS are described in the next section.

Human Mental Constructs Represented – Activity: Activities define the behavior of the human agent. They are a part of the simulated operator's procedural knowledge contained in the Updateable World Representation (UWR) and form the backbone of the simulation. Activities are scheduled or queued before being executed. The human agent's scheduling behavior is based on Wicken's multiple resource theory (1999), where parallel activities can be performed, if resources (visual, auditory, cognitive and psychomotor) are available. If sufficient resource is not available for concurrent performance (using a simple additive model) then, these activities can be interrupted by a higher priority activity, and later resumed.

Memory: The physical world is sampled regularly by the agent's perceptual and attention resources and the sampled data is stored in the UWR. Working Memory (WM) has been implemented in MIDAS based on postulates described by Baddeley and Hitch (1974). WM consists of a central control processor (with limited capacity), a "phonological loop" (temporary storage of speech-based information), and a "visuo-spatial scratch pad" (temporary storage of spatial information). Long Term Memory (LTM) is composed of both episodic and procedural archival structures. Both WM and LTM are susceptible to decay of information stored therein, caused by the passage of time since the information was last accessed, and to capacity overloads.

Goal Definition. Goals for every specific condition in the simulation world have to be defined. A goal is a statement of conditions (defined as "perceivable states of the simulation") that are to be met in its satisfaction. A goal is satisfied by decomposing the goal into "sub goals and activities"—these are defined by subsumption principles to provide a set of basic activities through which the human operator model interacts with other human agents in the simulation as well as with the equipment in the simulation. Basic (or leaf level) activities are defined as the point at which the action of the agents of the simulation are effected through an interface with the simulation world.

Activities. A set of goals and sub goals are decomposed into component parts that use an elementary information processing step in the human model and specific equipment. Activities are allocated resource loads associated with the elementary information processing aspects of human models. These loads are assigned in terms of visual, auditory, cognitive, and motor (VACM) requirements for an action to be performed. Activities also have duration estimates—and distributional variation around those estimates— used for scheduling the intended performance time of an activity. Each activity has a priority assignment that is inherited from the goal associated with it. They also have interruption specification (whether or not they can be interrupted once begun) and resumption specifications if interruptible. Activities and goals are the processes by which the human operator model interacts with the simulation world. Activities also have specification in their "goal decomposition" methods that assign logical processes (Boolean logic) to a task-type (e.g. activities that can be performed in a parallel fashion, activities that must be performed sequentially, do-while background or loop activities etc.).

Operators and Agents. Each operator (human or artifact) have software methods associated with it that track its interaction with other agents in the world. These "biographers" are used to collect the data of the transaction for agents in the simulation world. Human Operator Agents have several unique characteristics that are important to the functioning of the simulation.

Scheduler. The human agent has a scheduler that attempts to schedule activities for the human agent at each schedule cycle. As described earlier, he scheduler assumes that concurrent performance is desired unless otherwise specified. It attempts to schedule all activities possible in a given time frame until the human resource limit is reached. Priority determines which activities are scheduled first. Activities of the same priority are scheduled by a probabilistic coin flip.

Air MIDAS is able to provide a variety of outputs, e.g., operator workload, task performance timelines, and order of task completion, depending on the level of detail of model construction. The method section defines how the various characteristics of the concept were implemented in the model and data collected for the same.

Method

An informal cognitive walkthrough of the AAC was undertaken with the SMEs (retired ATC and AAC's concept developer) and comparisons drawn between this concept and current day operations. The concept was examined from the perspective of an air traffic controller working in the enlarged sector with traffic loads approximately double to those of current operations. At this preliminary stage of analysis, several simplifying assumptions were made to provide an initial implementation of the system that could be modified for further analyses.

Assumptions. It was assumed that there will be a single controller position (r-side) interacting with AACS with decision aids being provided through the set of tools usually used by the radar controller. All aircraft in the simulation are assumed to be equipped for AAC operation except for the current day operations condition.

The scenario provided that a single aircraft will transition from AACS through T-SAFE to ATC's manual control, and the controller will handle that particular aircraft until it leaves the sector. During the failure mode condition, the T-SAFE system provides a three-minute conflict free trajectory in the transition out of AAC mode and other controller tools provide support after that point. The T-SAFE system is not used after the transition. Communications are assumed to occur primarily through data link coordination between AAC and aircraft (and between controller and aircraft in standard/current day operations). Three scenarios were encoded - Current operations, AAC operations, and T-SAFE operations. The next section details the procedures for each kind of operation.

Procedure Definition

Three different kinds of procedures were simulated that focused on the en-route phase of the flight in this research effort. Also the role of r-side Air Traffic Controller was of prime focus. In the system, three kinds of agents were represented- Air Traffic Controller which was a Symbolic Operator Model (SOM) agent, AACS, and T-SAFE were represented as equipment agents. The main difference between a SOM agent and an equipment agent is that a SOM agent performs tasks specified by the task scheduler that uses estimates of human resources and priorities to schedule tasks, whereas the equipment agent has no such task scheduler.

Current Operations. The standard operations for the controller monitoring traffic, detecting conflicts, and resolving conflicts were encoded. Handoff procedures similar to the current day operations, where the controller via automation flashes the aircraft to be handed off on the ATC display. The controller in adjacent sector notes the flashing aircraft, prepares for handoff, and accepts the handoff. The previous sector controller notes that handoff has been accepted and accordingly requests aircraft to switch frequency to the next controller. Similarly conflict detection and resolution algorithms were formulated for this condition.

*AAC Operations***.** A conflict free scenario was encoded to depict conflict free trajectories created by the AACS. All aircraft in the simulation were considered equipped and under AACS control with just one controller handling them. The task of the controller was primarily to monitor traffic. Handoffs between sectors were handled by the automation (AACS).

*T-SAFE Operations***.** The operations using T-SAFE were procedures for transition between automated and manual / standard operations. This occurs when an equipped aircraft due to some reason (failure) changes status to unequipped aircraft. T-SAFE computes a 3 minutes conflict free trajectory for the failed aircraft before handing-off the aircraft to the ATC agent. After that the controller assumes manual control of the aircraft, T-SAFE has no role to play for that aircraft.

Procedures for the AAC & T-SAFE operations scenario and sequence of activities include the human operator agent monitoring the state of the airspace as a part of his/her standard goal of maintaining situation awareness. This monitoring for Situation Awareness (SA) goal is a background "do-while" activity. If an AAC T-SAFE alert is heard or seen, the operator agent ceases the standard SA scan and begins the goal of preparing to accept hand-off from T-SAFE. This handover occurs because the T-SAFE goal is a higher priority than the monitoring goal, and when interrupted the scheduler finds that the resource demand for the T-SAFE set of activities is high, therefore the activities cannot be performed in parallel. As will be discussed in the results section of this report, due to memory limits the information that the controller agent may have about the airspace into which the transition is occurring may be deficit. So a series of information seeking activities are initiated.

Airspace Definition

The airspace used to test the procedures was sector 47 and 49 in the Cleveland (ZOB) Center. The two sectors were combined to create “super sector” as described in the AAC concept. Only four major routes were simulated in the combined sector, and they all intersected close to the Cleveland airport at Dryer (DJB). The four routes represent traffic flows in the north-south, east-west, northeast to southwest and southeast to northwest directions and vice versa.

Aircraft and their trajectories were selected from the ETMS data for August 28, 2002. Aircraft that were enroute for the combined-sector were selected for the simulation, which meant that arrivals and departures out of the combined-sector were excluded for this phase of the simulation. The number of aircraft in the current ops was half the number of aircraft in the AAC and T-SAFE mode. The AAC and T-SAFE scenarios had 32 aircraft whereas the current operations had 15 aircraft.

Model Caveats and Constraints

As noted earlier and summarized here, several constraints need to be kept in mind while interpreting the results of this simulation. First, MIDAS does not have a complete efficiency and flow referenced set of air traffic procedures. So comparison of the performance of the model as a comment on the expected utility of the AAC to control and manage traffic is not appropriate.

Second, other support tools that would presumably be available to the controller have not been modeled, to assist the ATC in the management of traffic in the transition from T-SAFE.

Third, while the traffic sample is realistic (being taken from ETMS data files) however, there are no weather or other anomalous events to engage the controller even when AAC operation is nominal.

Fourth, the model lacks the implementation of the “critical maneuver” support techniques that are postulated to be part of T-SAFE, so the relative contribution of these to traffic control is not predicted.

Results and Discussion

The simulation focused on understanding the impact of the advanced airspace concept versus current operations on procedure, with respect to changes in workload, and status of goal completion for the controller agent.

Workload

The advanced airspace concept argues that a limiting factor to the capacity in the en route National Airspace is the workload experienced by the en route air traffic controller. Thus the analysis of estimated workload was performed for three different operational scenarios- Current operations, AAC operations, and T-SAFE operations. It was hypothesized that the estimated average workload for current operations would be the highest; it would decrease under AAC operations, and again increase for the T-SAFE operations. It is interesting to note that the workload estimated for the current operations and T-SAFE is the same (Figure 1), although there is a big difference in the number of aircraft. T-SAFE had 32 equipped aircraft with only one unequipped (due to unspecified failure), where as there were only 15 aircraft under ATC’s manual control in the current operation condition. Thus monitoring just one unequipped aircraft along with 31 equipped aircraft forces the controller to operate within narrow boundaries that increases the controller’s workload.

Several other studies (Corker, Gore, Flemming and Lane, 2001 or Jara & Corker, 2002) on mixed equipage operations have found that mixed operations can be challenging to the controllers. Corker, Gore, Flemming & Lane (2001) studied the impact of mixed equipage by changing the percentage of aircraft in free flight (standard ops, direct routing, 20% aircraft in free flight and 80% aircraft in free flight), and found that controllers reported highest subjective workload for the condition with 80% free flight aircraft.

Jara & Corker (2002) conducted a part task simulation study on controllers with varied control modes and also manipulated the presence or absence of a secondary task. The conditions designed with respect to the secondary task were referred to as shared and traded supervisory control. Shared control is the performance of a task by a human operator with the concurrent assistance of automation. In the shared supervisory control, the specialists monitored the airspace with no distractions from a secondary task. The traded condition represents a control style where either automation or human is in complete control, thus a secondary task was introduced in this condition. The researchers also found that controllers experienced significantly higher workload in the traded condition. Traded condition is somewhat analogous to the T-SAFE condition because it involves an unequipped aircraft, which is equivalent to the secondary task in the part task simulation.

In looking at the averaged workload in some detail (Figure 2), we have selected three sequences associated with conflict detection and resolution in current operations or baseline, T-SAFE handoff and normal AAC operations. These are represented in the following figures. The time scale for these graphs is the completion time for each activity (roughly a time line or an event line).

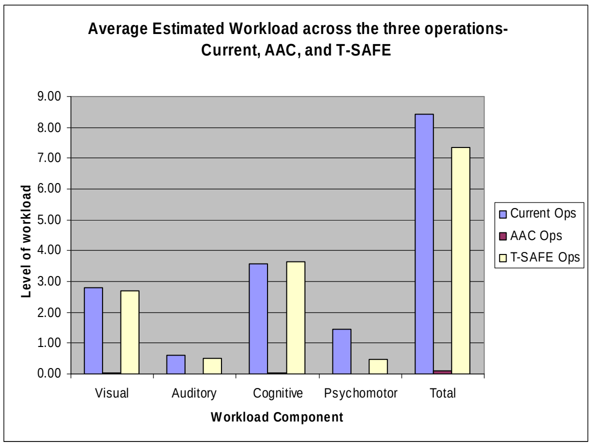


Figure 1. *Average Estimated Workload in the three operations – Current Day, AAC normal and T-SAFE. The workload scale ranges from 0 to 7 on every workload component.*

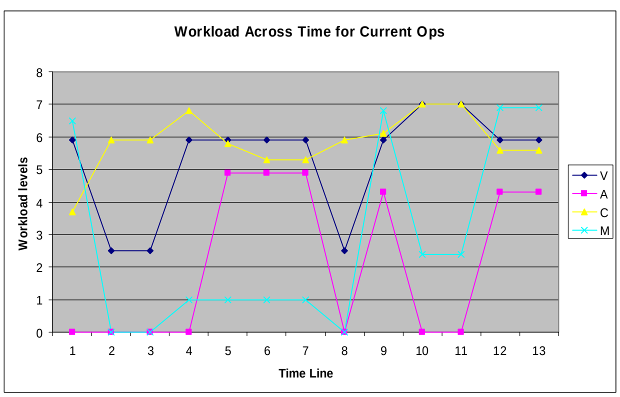


Figure 2. *Workload in visual, auditory, cognitive and motor (V,A,C,M) terms for activities associated with managing and separating approximately 15 aircraft in a sector in the current operations condition.*

Figure 3 shows the workload associated with managing a single aircraft in transition through T-Safe to manual control while managing approximately 31 other aircraft. It is clear from examining this workload trace that the predominant workload in this process is that associated with regaining awareness of the airspace into which the transitioning aircraft enters. This update is based on the requirement in the model to have current information in working memory to carry out the goals associated with aircraft conflict detection and resolution. Cognitive and visual load is high associated with tasks required for situation update.

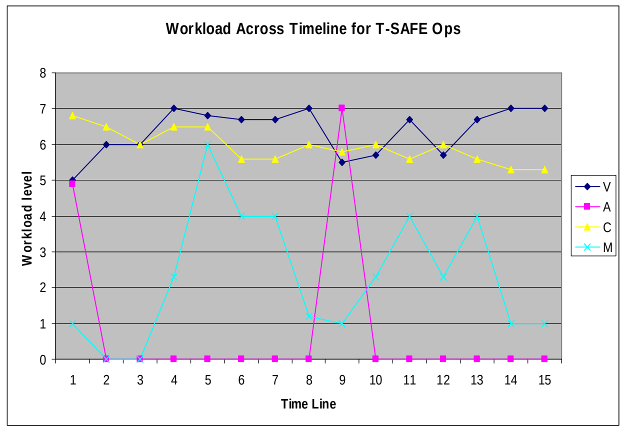


Figure 3. *Workload components (V,A,C,M) in the T-SAFE operations with one unequipped and 31 equipped aircraft.*

Status of Goals

The number of goals completed is an index of "how busy" the controller is, and number of aborted goals provides a sense of resource constraints experienced by the controller agent. The status of goals (completed and aborted) has a trend similar (Figure 4) to the average workload data. The number of goals completed is the highest in the current day operation because the controller is manually managing the traffic. It is interesting to note that number of goals handled under T-SAFE operations is high, where the controller agent is handling only one unequipped aircraft with rest of the traffic being handled by automation. These data are similar to the communication time data explored in the study by Corker et al. (2001). They found that although the controllers reported highest workload in the highest mixed equipage condition (80% free flight), they actually experienced highest communication load in the 20% free flight condition. Thus it is possible that increase in communication with a small percentage of mixed equipage (one failed or unequipped aircraft in the T-SAFE operations) can increase the number of tasks (mostly communications tasks) handled by the controller.

In terms of number of aborted goals, Figure 4 shows that about equal number of tasks/goals were aborted in the current and T-SAFE operations. Task shedding is a common response to information overload. Aborted tasks correspond to slips identified by Reason and Mycielska (1982) as causes of errors. They explain that slips occur when well formed plans are poorly executed due to omission of tasks, or intrusion of unwanted tasks.

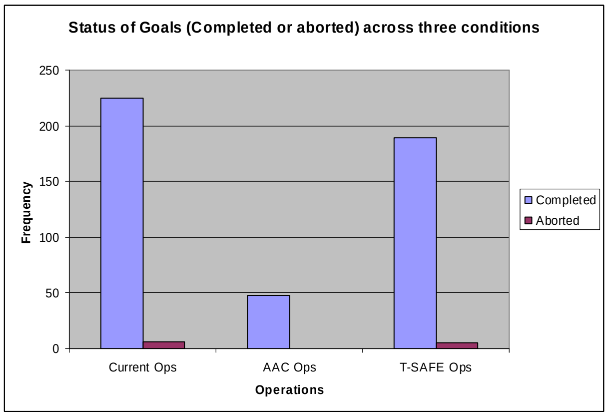


Figure 4. *Number of Goals completed and aborted across three conditions* **-** *Current Day, AAC normal and T-SAFE*

Conclusion

The purpose of this research effort was to model & analyze the current state of development and definition of the Advanced Airspace Concept Operations and using human-system performance model (Air MIDAS) to probe its impact on air traffic controller behavior. In order to examine the capacity benefits of AAC, current day standard operations, but with high traffic load were also modeled. It is clear that under normal conditions AAC operations significantly reduce workload for the controller. In this simulation twice the airspace and twice the traffic were handled in AAC operations by a single controller as compared with a controller team in current operations. However, one unequipped aircraft handled under T-SAFE operations can potentially increase workload to levels that approximate current day high load operations. The constraints, under which the current analysis was performed, have been explicitly stated. These constraints on assumed equipage and procedure can be relaxed to explore more refined representations of the operational concept. Future recommendations for research include examining any vigilance decrements under AAC operational mode due to extremely low levels of workload. Another recommendation would be to test more than one unequipped aircraft in the T-SAFE operational mode. It will be interesting to investigate the impact of the position of the failed or unequipped aircraft on workload. The position of the failure of aircraft will determine the cognitive resources required by the controller-agent to reconstruct her situation awareness.

References

Corker, K. & Smith, B. (1992). An architecture and model for cognitive engineering simulation analysis: Application to advanced aviation analysis. Presented at AIAA Conference on Computing in Aerospace, San Diego, CA.

Corker, K. (2000). Cognitive Models & Control: Human & System Dynamics in Advanced Airspace Operations. (2000) in N. Sarter and R. Amalberti (Eds.) Cognitive Engineering in the Aviation Domain. Lawrence Earlbaum Associates, New Jersey.

Corker, K., Gore, B.F., Fleming, K. & Lane, J. (2001). Air-ground Integration Dynamics in Exchange of Information for Control. Transportation Analysis: New Concepts and Methods in Air Traffic Management. (Bianco, Dell'Olmo and Odoni (Eds.) Springer Verlag.

Erzberger, H. (2001) The Automated Airspace Concept. FAA/Eurocontrol Workshop, Albequerque, New Mexico, December.

Jara, D., & Corker, K. (2002). Supervisory control: Air traffic control specialist performance in shared versus traded states. In Proceedings of the 21st Digital Avionics Systems Conference, Irvine, CA.

Reason, J. and Mycielska, K. (1982). Absent Minded? The Psychology of Mental lapses and Everyday Errors. Englewood Cliffs, NJ: Prentice-Hall.

Wickens, C.D. (1999). Attention in perception and display space. In J. Hollands & C. Wickens (eds.), Engineering psychology and human performance (3rd edition). N.J.: Prentice Hall.

Baddeley, A. D. and Hitch, G. J. (1974). Working Memory. In G. Bower, ed. Advances in Learning and Motivation. Vol. 8, New York: Academic Press, pp. 47-90.

SUPPORTING PILOTS IN RECOVERING TRAJECTORIES WITH TUNNEL-IN-THE-SKY DISPLAYS

Ernst J.G. Verschragen, **Max Mulder**, **M.M. (René) van Paassen**

Delft University of Technology, Faculty of Aerospace Engineering,
Control and Simulation Division
Delft, the Netherlands

Tunnel-in-the-sky displays have shown great potential for reducing pilot workload and navigation error. Although it is a well-evaluated concept, only little research has been conducted on situations in which the pilot has (deliberately or not) flown outside of the tunnel. This paper describes the experimental evaluation of various alternatives to support pilots in recapturing their nominal trajectory. The concepts studied include the use of guiding arrows, path deviation indicators, a symbol representing the tunnel and a "return tunnel". Results from a pilot-in-the-loop experiment indicate that a "return tunnel" performed best on situation awareness and workload aspects and that most pilots participating in the experiment had a general preference for this support concept.

Introduction

Tunnel displays may enable aircraft to closely follow intricate trajectories as a means for improving air traffic management efficiency, and for meeting noise abatement concerns (e.g. Grunwald, 1984, Mulder, 1999). While studies to date have demonstrated the potential benefits of tunnel displays, only little research has been conducted on situations in which the pilot has flown outside of the tunnel. The tasks of determining the aircraft's position in relation to the tunnel, creating a mental image of the situation, and generating a recovery path to intercept the original trajectory can be very demanding at times. Therefore, track-recovery support (TRS) is necessary to enable the pilot to reacquire the planned trajectory.

Several papers indeed state that TRS is an "issue" for TIS displays (Beringer, 1999, Theunissen, Rademaker & Etherington 2002, Newman, 2003, Newman & Mulder, 2003). Besides scenarios in which the original trajectory is abandoned for some reason, the TRS symbology is also applicable in a transition from a flight phase with low (or less) precision guidance to a flight phase with high (or more) precision guidance. In this case, the TRS symbology directs the pilot to the beginning of the precision path. Theunissen et al. addressed the transition from conventional navigation modes to guidance along a complex, tightly constrained path. They considered two path-intercept concepts: a flight path predictor with a reference marker (directional guidance) and a 3-D intercept-path towards the precision path. From an initial evaluation of the flight path predictor with target marker concept, they concluded that the task of the pilot using this display is similar to using a predictor reference box when flying on the fixed path. They did not test the 3-D intercept-path concept, arguing that there is no difference between flying a 3-D intercept-path or flying the original precision path. Williams (2000) tested how well pilots were able to acquire a pathway in the sky with several types of guidance. It was found that a follow-me airplane yielded best performance over a flight predictor and the no-guidance condition.

This work studies the return maneuver as a whole. The starting point is an aircraft that has strayed from the nominal trajectory: it has an arbitrary cross track error and track-angle, and visual contact with the tunnel may be lost. Four TRS concepts will be described that assist the pilot in finding and recapturing the precision path. These concepts were selected after a theoretical investigation of initially eight concepts (Verschragen, 2004). The results of an experimental evaluation in which these concepts are compared are described.

Approach & Preliminaries

The TRS concepts are intended for path finding as well as path intercepting support. Supporting pilots in recovering the tunnel can be done in two ways:

1. Provide the pilot with information on the <u>status</u> of the aircraft: i.e. the position and track in relation to the tunnel. The status information enhances the situation awareness of the pilot, enabling him to form a better recovery strategy.
2. Provide the pilot with <u>guidance</u> information; the pilot follows the guidance commands given to him by the display. This approach relieves the pilot of the task of forming a recovery strategy.
3.

In case of guidance information, a recovery algorithm is needed that computes either a recovery path or, in case of directional guidance, a commanded track-angle. In case of status information, the pilot determines a recovery path himself. Either way, the information for track-recovery support has to be presented through display symbology. The next section discusses four TRS concepts. Each concept will be explained on the basis of the intelligence needed for the concept and the way the information is presented on the display. We assumed the following:

- The TRS elements are presented only on the Primary Flight Display.
- The Navigation Display is not considered, as it could hide differences between the concepts.
- The nominal trajectory is straight.
- Only the horizontal plane is considered.
- The aircraft velocity is constant.
- The effects of wind and turbulence are neglected.

TRS Concepts

Four TRS concepts will be discussed in this paper, for a more detailed analysis including other concepts the reader is referred to (Verschragen, 2004).

Arrows Concept (AR)

The arrows concept (AR) provides directional *guidance*: the pilot is instructed to fly in a certain direction (Newman, 2003). The pilot is presented with the track (and flight-path) angle error of the aircraft with respect to the desired track (flight path) angle, Figure 1. The size of the arrows is related to the magnitude of the errors represented by them. When the aircraft flies at the desired track (or flight path) angle, the arrows disappear.

The intelligence behind the arrows is in analogy with the procedure of flying to a VOR beacon. Three areas are defined (parallel to the tunnel), each with a different commanded track-angle. Figure 2 shows these three areas and the trajectory shape intended by the arrows.

Tunnel Symbol Concept (TS)

The tunnel symbol (TS) concept is a *status* information concept: it provides the pilot with the difference in track angle (and vertical off-set) with respect to the tunnel, Figure 3. As compared to the AR concept, no “error” is shown, but a difference in track-angle. A pilot derives his recovery strategy.

The intelligence behind the display derives the track-angle difference between aircraft and tunnel (and vertical off-set). No recovery path or direction is generated. Rather, the algorithm determines whether the aircraft is flying in the correct tunnel-direction and also whether it is flying towards or away from the tunnel. Then, the track-angle error is shown by a hatched plane that rotates about its vertical center axis, Figure 4. The vertical position deviation of the aircraft with respect to the tunnel is clarified by the vertical position of the tunnel symbol on the display. When a change in flying direction in relation to the original tunnel occurs, the tunnel symbol flips from the left to the right side (or vice versa). The cross-track error is shown qualitatively by the scaling of the tunnel symbol; at a certain size of the cross-track error, the symbol will not become smaller if the cross-track error becomes larger, this to keep the tunnel symbol readable. The symbol is shown in green if the direction of flight is within +90 or -90 degrees of the tunnel-direction, otherwise it is shown in red. Furthermore, the symbol is fully drawn if the pilot is flying towards the tunnel and dotted if the pilot is flying away from the tunnel.

Return Tunnel Concept (RT)

The return tunnel (RT) concept is a path-based *guidance* display: a 3-D path leading back to the tunnel is presented to the pilot. For the RT concept implementation an elaborate algorithm was defined (Verschragen, 2004). The “return tunnel” only differs from the original tunnel by its green color, Figure 5. The return tunnel is generated when the pilot presses a button. It does not move along with the aircraft but remains a static object in the world and approaches the original tunnel with an intercept angle of 30°.

Deviation Indicators Concept (DI)

The deviation indicators (DI) concept offers only *status* information; the pilot is provided with the horizontal and vertical deviations from the planned path. Also the rate at which these deviations increase or decrease is given to the pilot, which indirectly informs the pilot about the difference in track between the aircraft and the tunnel.

The deviation indicators consist of one horizontal and one vertical scale that indicate the position of the aircraft in relation to the planned path., Figure 6. In analogy with the “follow-the-needle” principle, the deviation indicators scale centers represent the own aircraft. A moving square on the scale shows where the original path is located. The indicators are extended with a yellow trend vector that indicates the velocity at which the cross-track error changes. The scales are linear and show deviations from -1000m to +1000m between the tunnel and the aircraft for both dimensions, small lines indicating another 250m.

Experimental Evaluation

The goal of the experiment was to evaluate the effects of the four TRS concepts on pilot performance, workload and situation awareness.

Subjects and instructions. Nine experienced professional airline pilots participated in the experiment. They were instructed to capture the tunnel in the way they thought best.

Apparatus. The experiment was run on a desktop computer in a noise-free room. The pilot controlled the aircraft motion with a joystick. With the RT concept a button could be pressed to generate the return tunnel. The 17'' computer screen showed a generic "tunnel-in-the-sky" Primary Flight Display extended with one of the experiment concepts.

The aircraft model. A linear model of a Cessna Citation 500 was used, trimmed for a speed of 77m/s. No wind or turbulence model was simulated.

Independent variables. Two independent variables were defined: the TRS concept (4 levels), and the experiment scenario (6 levels). The scenarios are shown in Figure 7. Scenarios 1 and 2 are considered more difficult than the others, because here the tunnel is not visible at the start of the runs. Vertical deviations from the nominal path are limited. Scenarios 3 and 4 have an initial vertical off-set compared to the tunnel of 100m (above the tunnel). In all other scenarios, no vertical off-set is used.

Experiment design and procedure. A full-factorial within-subjects design was used, yielding 24 conditions. The subjects first received instructions on how the TRS concepts worked and had a chance to fly them as many runs as needed to understand the concept. After the learning phase, the pilot flew the 24 conditions in a randomized order. After the experiment, a pilot questionnaire was handed out, querying pilots on performance, workload and situation awareness aspects of the TRS concepts.

Dependent Measures

The following dependent measures were defined:

- First-turn errors: maneuvering in the wrong direction at the start of the run. The RT return trajectory of is taken as the "correct" trajectory.
- Spread of mean cross-track error: a measure to determine the diversity of the return routes.
- Number of stick inputs: several counters were defined to separate the larger stick deflections from the smaller ones. The counted number of deflections per run is divided by the run time.
- Total return time: the time it took pilots to guide the aircraft back into the tunnel.
- Maximum cross-track error: the maximum distance to the reference path.
- Constant track interval: the longest "straight" segment (track-angle error smaller than -1°/+1°) is measured relative to the total run time.
- Time to final atd: the total time necessary to not only return to the original trajectory, but also to reach a certain atd at the original track.

The last performance measure is used to determine which concept allows the most *time-efficient* returns. To be able to compare different runs, the run with the farthest atd at the moment of intercept, atd_{final}, is selected. For all other runs of the same condition, the time necessary to reach this atd_{final} was calculated :

$$T_{to\ finalatd} = T + \frac{atd_{final} - atd_{intercept}}{V_{intercept}},$$

in which T stands for total return time, $atd_{intercept}$ is the atd at intercept of the particular run and $V_{intercept}$ is the velocity of the aircraft at the moment of intercept.

Experiment hypotheses. First, it was hypothesized that the amount of first-turn errors would be lowest with the RT and AR displays, and the spread in mean cross-track error would be smallest with the RT display. Second, it was hypothesized that RT and DI would yield in the lowest and highest workload, respectively. It was expected that the RT display would yield more control activity than the other displays, because the pilot would try to stay inside of the return tunnel. Third, as far as performance is concerned, it was hypothesized that the RT display leads to the longest return times. The RT algorithms create a return route with a small intercept angle, and therefore gradually reducing the cross track error. For the AR concept, the return times were expected to be smallest, because the pilot is directed perpendicularly to the original tunnel. With the other two concepts the return times are hypothesized to lie in-between.

Fourth, it was hypothesized that the RT concept would provide the most efficient returns; the RT algorithms were designed to minimize time-loss incurred by the out-of-tunnel incident. The maximum cross-track errors were expected to be largest for the RT display, because the RT algorithms create a return route that is not optimized for minimizing position errors, while with other concepts, the pilot can apply his own preference. Finally, it was hypothesized that RT leads to the longest sections of constant track-angle, again as it is inherent to the RT algorithm.

Results

A full-factorial ANOVA was conducted on most dependent measures. Some data were defined as "counters" (e.g., the number of f-t errors and the number of stick deflections), data that is not necessarily normally distributed. Here, the non-parametric Friedman test was used. If it revealed a significant effect, Wilcoxon tests were executed to compare each of the displays separately. Figure 8 shows the means and the 95\% confidence limits of some of the main dependent variables, for one difficult (1) and one easy scenario (3 or 5).

Table 1. *Number of first-turn errors per condition.*

Scenario	TS	AR	RT	DI	total
1	5	0	1	2	8
2	4	0	0	0	4
3	2	2	1	2	7
4	2	0	0	0	2
5	3	2	1	1	7
6	2	0	0	1	3
total	18	4	3	6	31

First turn errors. The amount of first-turn errors was significantly influenced by the display type (χ^2_3 =13.57, p<0.01). Wilcoxon tests revealed that the TS concept leads to more first-turn errors than the other three concepts. Furthermore, with the RT concept, less first-turn errors are flown than with the DI concept (Table 1).

Diversity return routes. Figure 8 shows that for scenario 1 (considered difficult) the spread in the mean cross-track error is largest for the TS display followed by the DI display. The RT and AR displays show the smallest spread. Scenario 5, however (considered simple), shows equally large spreads in mean cross-track error for all concepts. Thus, in hard scenarios it becomes clear that RT and AR allow the pilot to fly a more precise route than the tunnel symbol and deviation indicators displays. In simple scenarios this effect is not (less) visible.

Control activity. The counters show a distinction between large and small aileron stick deflections. Only a marginal significant influence is found of the display format. Wilcoxon tests revealed that for large stick deflections TS is outperformed by the RT (α=0.05, p= 0.0256) and DI (α=0.05, p=0.0629) concepts. For small deflections, Wilcoxon tests revealed that DI was outperformed by AR (α=0.05, p=0.0650) and TS (α=0.05, p=0.0830) concepts. The AR concept performs better than the RT concept (α=0.05, p=0.0830). For small stick deflections the RT concept shows the highest control activity. These are due to the tracking of the return tunnel.

Fastest return. A significant effect on the total return time was found for both the display ($F_{3,21}$=21.587, p<0.01) and scenario ($F_{5,35}$=42.950, p<0.01) Furthermore, a significant 2-way interaction was found $F_{15,105}$=5.947 (p< 0.01). A Post-Hoc analysis (SNK, α=0.05) revealed that the RT display performed worse than the other three displays.

Most efficient return. The time to final atd was significantly influenced by the display type ($F_{3,21}$=4.131; p=0.019). The scenario significantly effected the time to final atd ($F_{5,35}$=91.733, p< 0.01). A Post-Hoc analysis (SNK, α=0.05) revealed that the RT display was outperformed by the AR display.

Minimizing position errors. The maximum cross-track error is significantly influenced by the display format ($F_{3,21}$=13.462, p<0.01) as well as scenario ($F_{5,35}$=27.095, p<0.01). The 2-way interaction was significant ($F_{15,105}$=7.407, p<0.01) as well. Post-Hoc analyses (SNK, α=0.05) for display revealed three different groups. The TS concept leads to the smallest maximum cross-track errors, followed by the AR and DI concepts in the second group. The RT display yields the largest errors.

Stable return maneuvering. Display ($F_{3,21}$=31.056, p<0.01) and scenario ($F_{5,35}$=11.367; p<0.01) significantly effected the variability on the return maneuvers. Again the interaction was significant ($F_{15,105}$=4.917; p< 0.01). Post-Hoc analysis (SNK, α = 0.05) for display revealed three groups. The RT concept performed best. A second group contains the TS and AR displays. The TS concept also forms a third group with the deviation indicators.

Questionnaire. The pilots indicated that the RT and AR concepts show which direction to steer to at the start of the maneuver in the most clear way. The size of the track-angle error is shown clearest with the RT and TS concepts, and the lateral position can be obtained easiest with the DI display. Pilots indicated that with the RT, their understanding of the flown trajectory after the run was best and that the RT display was the most intuitive. Subjects indicated that the RT concept improved situational awareness most.

Also, with the RT the capture maneuver costs least effort. It was considered the most comfortable and the DI concept the least comfortable display. Workload was found lowest for the RT concept, followed by the AR concept. Regarding performance, most pilots were of the opinion that the RT allows best performance in general. Overall, six of nine pilots preferred the RT concept.

Discussion

From the statistical and subjective results, it appears the RT display offers the best situational awareness; the amount of first-turn errors is smallest for RT display, indicating that the RT offers the clearest symbology. The diversity between the return routes were smallest for this display. Pilots indicated that the RT provided the most intuitive display and helped them best to understand the trajectory flown.

The RT display leads to the highest control activity for small stick deflections. This is caused by the tracking of the tunnel and not to intercepting the nominal trajectory. RT resulted in the lowest amount of large stick deflections, indicating that pilots felt comfortable with this display.

The AR display allows the most efficient return, while it was hypothesized that the RT display would perform better on this measure. This discrepancy is due to the design of the RT algorithm: with small deviations from and flying with large intercept angles towards the tunnel, the return trajectory will first cross the original tunnel before initiating a turn to final intercept of the planned path.

The TS display leads to the smallest position errors, but it also resulted in the steepest turns. This can be expected since if the aircraft is flying away from the tunnel initially, a steep turn (i.e. a smaller radius) will limit the maximum position errors.

The RT concept leads to the longest and least efficient returns with largest position errors, because the algorithms that produce these return tunnels were not designed to optimize these measures for performance. The return algorithms generate stable return trajectories that gradually approach the original path. Therefore, the RT display leads to longest intervals of constant track-angle. Furthermore, the pilots indicated that they felt they performed best with the RT display.

Conclusions

The objective of this work was to evaluate four track-recovery support concepts. It was found that the return tunnel concept (RT) offers the best situational awareness and the lowest control activity in terms of large stick deflections. It also led to the largest return times and position errors, but this is inherent to the algorithms that calculate the return tunnels. Obviously, because pilots relied on the tunnel guidance, it resulted in minimal variations in return maneuvers. And because pilots tried to accurately fly the tunnel return trajectory, the highest number of small stick deflections was found with this concept. The RT concept was preferred by most pilots. Performance with the RT can be enhanced by modifying the algorithms that calculate the return trajectories. Control activity can be reduced by optimizing the (return) tunnel dimensions.

For future research, some extensions should be made to the experiment design. First, the reference track should include one or more curved sections. This will imply a redesign of the return trajectory algorithms. Thrust settings should be incorporated as well, resulting in a variable speed and therefore a variable radius of turn. Most importantly, a navigation display should be taken along in the experiment, which contributes significantly to the situation awareness. The role for the track-recovery support display will then shift more to supporting the pilot in performing a smooth intercept with the reference trajectory.

It is recommended that future experiments include high-workload situations, in which the pilot has to divide his attention between different tasks. An intuitive display will pay off in these situations, because processing information will demand less of the pilot. It is hypothesized that the RT concept will outperform all others under these circumstances.

References

Beringer, D.B. (1999). Flight Command Guidance Indicators for Pathway Displays: My Way or the Highway? *Proceedings of the 10th International Symposium on Aviation Psychology*.

Grunwald, A.J. (1984). Tunnel Display for Four-dimensional Fixed Wing Aircraft Approaches. *Journal of Guidance and Control, 19*(2), 370-377.

Mulder, M. (1999). *Cybernetics of Tunnel-in-the-Sky Displays*, PhD dissertation, Faculty of Aerospace Engineering, Delft University of Technology.

Newman, R.J. (2003). Advanced Display Certification Issues. Technical Report, University of Iowa, 2003.

Newman, R.J. & Mulder, M. (2003). Pathway Displays: A Literature Review. *Proceedings of the 22nd Digital Avionics Systems Conference (DASC-22), 9D.6-1 – 9D.6.10.*

Theunissen, E., Rademaker, R.M. & Etherington, T.J. (2002). Path Intercept Functionality for Perspective Flight-path Displays. *Proceedings of the AIAA Modeling & Simulation Technologies Conference, Monterey (CA),* paper 2002-4788.

Verschragen, E.J.G. (2004). *Track-Recovery Support for Tunnel-in-the-Sky Displays*. Unpublished MSc thesis, Faculty of Aerospace Engineering, Delft University of Technology.

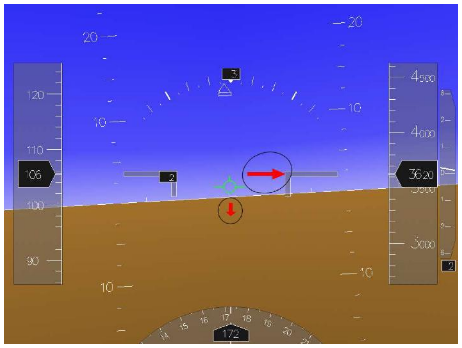


Figure 1. *The arrows concept AR (encircled).*

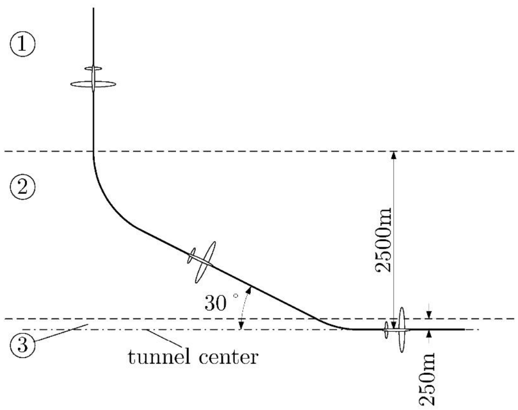


Figure 2. *The trajectory that is the result when following the arrows (AR) recovery concept.*

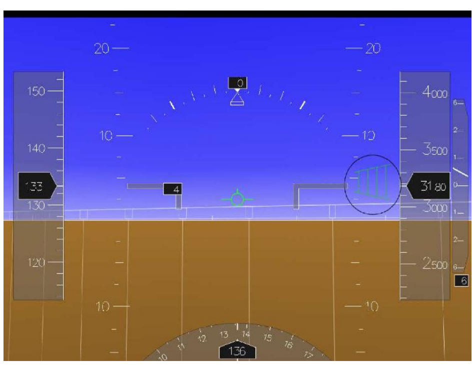


Figure 3. *The tunnel symbol concept TS (encircled).*

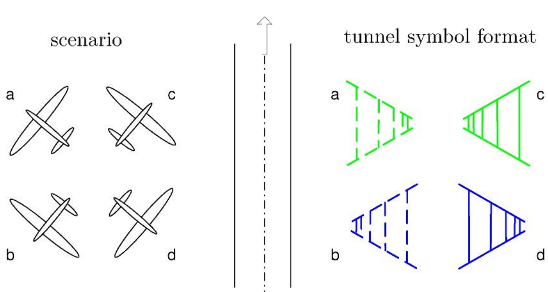


Figure 4. *Tunnel symbol color and line attributes.*

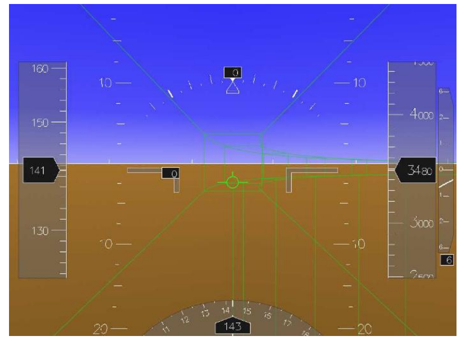


Figure 5. *The return tunnel (RT) concept.*

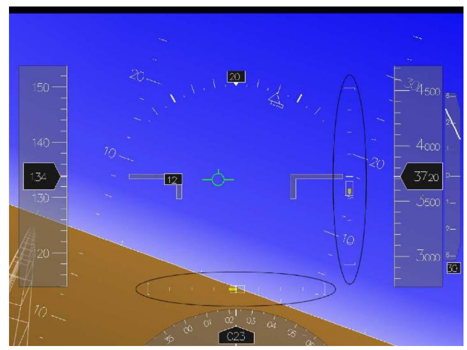


Figure 6. *The digital indicators (DI) concept.*

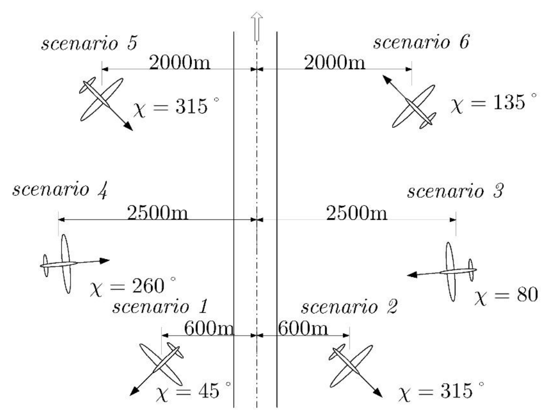


Figure 7. *Experiment scenarios definition.*

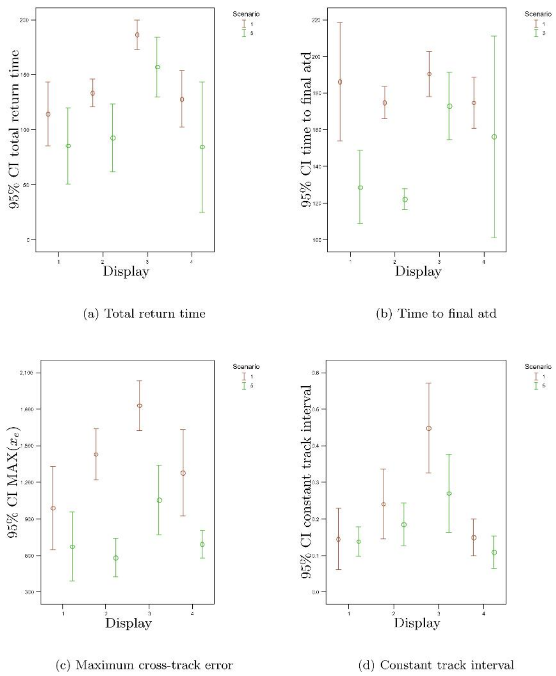


Figure 8. *Means and 95% confidence limits for the main dependent measures.*

DISTRIBUTED INFORMATION BEHAVIOR AMONG FLIGHT CREWS IN A SIMULATED ENVIRONMENT

Terry L. von Thaden, PhD
University of Illinois at Urbana Champaign
Champaign, IL

The purpose of this study was to assess concepts from Information Science to develop and initially validate a framework to study the information behavior of flight crews in the civil aviation domain. Distributed use of information within groups remains a weak link between actual information, the meaning given to information, and the sense made of the information. Principles from information science, psychology, and communication studies are used to analyze how flight crews in a simulated environment (fail to) make use of essential, safety critical information through analysis of the corresponding flight transcripts using a six-point Information Behavior Grid. The results of this research indicate differences in the way flight crews identify, gather and use information based on their performance level. This study discerns that high performing flight crews practice different information behaviors than low performing or accident involved flight crews. This work serves as a way to operationalize crew resource management through understanding the social practice of information structuring within the distributed collective practice of the flight crew. This work also serves as a tool to inform crew training and is applicable to other domains where work is supported through distributed collective practice.

Introduction

In aviation operations, flight crews must incorporate efficient and effective communication of essential safety-critical information to avoid accidents. Consistent, procedural responses to clearly defined situations are a normal part of conducting a flight, yet there are frequently indeterminate circumstances under which crews must use their personal judgment to negotiate meaning with members of their team to arrive at a solution. The process of how people employ sources and channels of information to satisfy a need is known as (human) information behavior (Wilson, 2000). This research explores the relationships between the distinct principles of crew information behavior, crew performance, and mission outcome to study the social construction of information in practice.

The Information Environment in Flight

Pilots must transform data from multiple interfaces into meaningful flight information. Given the safety critical nature of aviation operations, pilots must incorporate efficient and effective communication of essential information to avoid accidents. Flightcrews are trained to employ consistent, procedural responses to clearly defined situations as a normal part of flight operations, yet there are frequently indeterminate circumstances under which crewmembers must use personal judgment and negotiate the meaning of their personal judgment with other members of the crew to arrive at a solution. As such, crewmembers become information resources in the larger, distributed environment.

On the flightdeck, information processed in concert with other crewmembers may be more robust than information processed by each individual, yet it requires social, organizational and even technological devices for the continued support of group information retrieval and effective information use in increasingly complex situations. Crewmembers may not be as effectively organized in their communication of information as they could be, leading to misinterpretation of information, and dangerous situations. According to Hutchins, "Social organizational factors often produce group properties that differ considerably from the properties of individuals" (Hutchins, 1996, p. xx). Efficient information retrieval and use relies on patterns of group size, individual interaction, interaction through time and distribution of knowledge. Thus the cognitive properties *between* group members depend on the character of the social organization of the group, rather than the cognitive properties of the individuals *in* groups. This social organization forms the basis from which to study distributed negotiation of information between group members on the flightdeck using principles from the domain of Information Science.

Information Practice

Information needs vary at different stages of a process. The distinction perhaps can be made between whether information is a thing or a process and whether information is objectively or socially constructed. Buckland (1991) notes that objects such as data and documents have the qualities of imparting knowledge or communicating information, serving as information *things*. An information *process* on the other hand, is concerned with the procedure of being informed, a change in knowledge, not just the discrete form of the information *thing*. While finding the *thing* is an end goal, users need to be able to get through the *process* and barriers to it, of deciphering just what is the necessary information *thing* and how to get it. To do this, a person employs their collection of individual abilities consisting of experience, knowledge, resources to gather information, use the information, and communicate this knowledge. This is what Marchionini (1995) designates as *personal information infrastructure*. According to Marchionini:

> "A personal information infrastructure is a collection of interrelating mental models for specific information systems; mental models for events, experiences, and domains of knowledge; general cognitive skills (e.g., inferencing, recognizing salience) and specific cognitive skills related to organizing and accessing information (e.g., filing rules, reading); material resources such as information systems, money, and time; metacognitive resources for planning and monitoring thought and action; and attitudes toward information seeking and knowledge acquisition" (1995, p.11).

As people use information, they develop mental models of the skills needed to access information and understand how information is organized. When technology is brought into the information process, it can augment cognitive skills by assisting the user in finding and using information. Technology can also change the strategies users employ to acquire information, confusing or disorienting them, thus impacting their abilities and performance. Therefore, when interacting with information people learn to take advantage of what is easily available or understandable.

Information Behavior Grid

The Information Behavior Grid (IBG) (von Thaden, 2003, 2004) was developed using principles from Information Science, human factors science and communication studies (Wilson, 2000; Ellis, 1989,1993,1997; Choo, Detlor, and Turnbull, 2000) to measure distributed patterns of information needs, seeking and use among distributed groups. Applied to this research, the IBG is a tool to distinguish whether accident involved flight crews practice different patterns of distributed information behavior than those of non-accident involved flight crews. In the context of this study, it is not about measuring human error nor distinguishing the precise moment a decision is made, but rather to observe social information interaction in an attempt to measure distributed information practice and use of essential, safety critical information. This is accomplished through analyzing transcripts of crew interaction during simulated flight using the IBG.

Given the dynamics and training in aviation operations, information behavior may be understood as either passive/conditioned behaviors or active/formal behaviors. Pilots tend put information into practice two ways, they actively engage in a methodical, systematic, defined process of making sense of the environment, an almost feed-forward activity (although the process actively engages understanding past events to make sense), or they passively, casually survey the environment or their instruments to evaluate the environment, a more experiential, "seat-of-the-pants" endeavor. In other words, pilots tend to function informally or formally, looking *at* or looking *for* information. These distinctions of information behavior allow a general understanding of their work practice. Although these categorizations may lose some of the crews' intricate information strategies, the real need is to understand whether they base their information behavior solely on personal experience or formal methodology. This grid has been updated in place of the original model developed by von Thaden (2003) and von Thaden & Wiegmann (2004)(for a complete discussion see von Thaden, 2004). Figure 1 shows the layout of the Information Behavior Grid described below.

	Information Need	Information Seeking	Information Use
Exploration	Casual/ Conditioned Identification	Casual/ Conditioned Gathering	Casual/ Conditioned Use
Exploitation	Methodical Systematic Identification	Methodical Systematic Gathering	Methodical Systematic Use

Figure 1. *The Information Behavior Grid.*

In *Conditioned Identification* (CI) general areas of interest are passively viewed (scanned) using casual or informal means. There is no specific information need communicated but simple queries may be formulated or addressed on broad search areas. *Conditioned Gathering* (CG) may consist either of broadly sweeping varied resources to detect change signals and take advantage of easily accessible information or CG may consist of passively fixating on a limited area or instrument. In *Conditioned Use* (CU), information may be discovered serendipitously through passively browsing a number of different resources. CU may also entail passively or habitually acknowledging a change within narrow boundaries or using personal rather than technical criteria to arrive at a decision. In *Methodical Identification* (MI) general areas of interest or trends are actively recognized using practiced viewing patterns (schema). Specific detailed targets are actively sought or simple specific needs are updated and expanded through an ongoing search. *Methodical Gathering* (MG) of information involves actively browsing in preselected sources or instruments using prespecified protocols (methods/ procedures) to acquire information, such as attending to a checklist. MG also consists of active, ongoing measurement. *Methodical Use* (MU) of information entails actively increasing specific knowledge about areas of interest, relevance, or change. Relevant information is used for determining a specific course of action. MU also entails meticulous confirmation (verification) of information.

Method

The purpose of this study was to empirically evaluate the distributed information behavior of flight crews in the simulated environment. Specifically, advanced student pilots in a CRM course at the University of Illinois' Aviation Division (operating under FAR Part 141 approved curriculum guidelines) were observed participating in simulated flight exercises consisting of 5 distinct mission-based dynamic scenarios to various destinations around the United States. Over the course of the semester, pilots completed classroom assignments and learned the concepts relating to societal/cultural, industry, governmental regulatory agency, organizational, group, and individual influences on behavior and crew resource management. Simulated flights served as practical experience to learn the concepts of CRM. Laboratory and flight sections used a multi-engine Frasca 142 Flight Training Device (FTD), complete with dual instrumentation and controls simulating a Piper Seminole with Lycoming IO-360-A1H6, 180 hp engines and a maximum takeoff weight of 3800 lbs. Participants were familiar with the Piper Seminole and Frasca 142 FTD as they had completed multi-engine ratings in a previous semester at the University of Illinois using the same equipment.

Two students flew each mission together as a crew in normal, abnormal, and emergency situations to gain practical experience for working together as a team. Before flight simulations (i.e., missions) commenced, pilots were required to ensure the necessary documentation was aboard each simulated flight as would be required in actual operations including proper checklists, operating manuals, maps, charts, and any other equipment necessary to conduct the mission. Each mission consisted of a mission briefing, preflight planning, the simulated exercise, and debriefing. Pilots were given the necessary information to plan their route of flight and obtain necessary weather and advisory information. The missions, and their consequences, were simulated just as they would be in real world applications. Pilots were to fill out the required paperwork at the end of each mission. In the case of violations, incidents or accidents, pilots were required to fill out the necessary reporting forms at the end of each mission. Instructors acted as Air Traffic Control and Company Briefers when appropriate.

Twenty students were registered for the course. Participation was completely voluntary, no monies were paid to participants, no interventions occurred as a result of their videotaped sessions and there was no penalty for non-participation. Nineteen students agreed to allow use of their performance data in the simulated flights to be utilized in this study. Participants (n = 19) were on average 22.75 years of age with an average total flight time of 389.11 hours.

Pilots completed 5 different simulated scenarios over the course of the semester. Each pilot flew each mission as the pilot flying (PF) and the pilot not flying (PNF) the aircraft. In each case, the Pilot in Command (PIC) consisted of the PF or Captain of the mission and the Second in Command (SIC) consisted of the PNF, or First Officer of the mission. These pairings allowed the pilots to each have a turn acting as Captain and First Officer in each scenario. This was achieved by having one pilot act as Captain for the first leg of the flight, and then switch roles for the second, or return, leg of the flight.

The sessions were videotaped and transcribed resulting in 49 usable transcriptions. Restricted recordings of crew pairings not participating in the experiment, occasional problems with video recording equipment or the audio portion of the recording, or simply forgetting to turn on the equipment, resulted in 49 usable taped scenarios. Twenty-four total distinct crew pairings were captured in the recorded simulations for the nineteen pilots. Though the entire simulated laboratory session was recorded, the last 20 minutes of each mission (i.e. approach to landing phase) was used for the present analysis. The missions were evaluated for outcome (accident or non-accident), crew performance (high, average, low), and independently coded for crew information behavior using the IBG.

During the transcription accuracy check, each mission was evaluated for crew mission performance (high, average, low) by observing the interactions of crewmembers during each mission. This classification allowed for an analysis of the mission as well as to determine how crew performance may vary across the semester. The following criteria were used to assess mindful attention and heedful interaction (adapted from Weick and Roberts (1993) discussion of collective mind). Crews that displayed high professionalism, preparedness, and carried out heedful interactions the majority of the mission were categorized as High Performance (HP). Crews that displayed low professionalism, were not prepared for the mission, and were heedless in their interactions the majority of the mission, were categorized as Low Performance (LP). Crews that displayed neither superbly high nor excessively low performance were categorized as Average Performance (AP).

After the crew performance was assessed, each of the transcriptions was hand-coded by the researcher at the speech act level for instances of crew information behavior, blind to crew performance and to mission outcome. Each speech act was coded considering the PIC, SIC, and Instructor as part of the flight environment. Communication that could not be understood was recorded as non-codeable (NC), and that having no relevance to the flight mission was coded as Not Pertinent (NP). Nineteen speech acts were listed as non-codeable throughout the 49 tapes and not included in the analysis. Where appropriate, Instructor's communications were coded when they acted as briefers or controllers in the mission, as they represent part of the flightdeck's information environment. Five transcripts were randomly selected from the various stages of coding (2 early, 1 mid, and 2 later in the process). These transcripts were then re-coded by the same researcher without access to previous codings. An intra-rater reliability test was performed on 5 selections using percent agreement. Intra-rater reliability resulted in the acceptable score of 0.88, with no further reliability testing performed.

Results

For the 49 missions 11,869 observed information behaviors were coded, with an average of 242 behaviors per case across the scenarios. Figure 2 shows that combining the data for all missions, MU accounted for the highest percentage of information behavior at 32%, followed by MG at 19%, and CU at 18%, then CG at 12%, MI at 10%, CI at 7%, and NP at 2%. When viewed in the aggregate, information use is greater than information identification and gathering, and methodical information use is greater than conditioned information use. Information gathering is greater than information identification, but less than information use, and methodical information gathering is greater than conditioned information gathering. Information identification is less frequent than information gathering and information use, and methodical information identification is greater than conditioned information identification. Overall, conditioned information behaviors appear less frequent than their methodical counterpart. Non-pertinent information behavior occupies the least percentage of behavior overall.

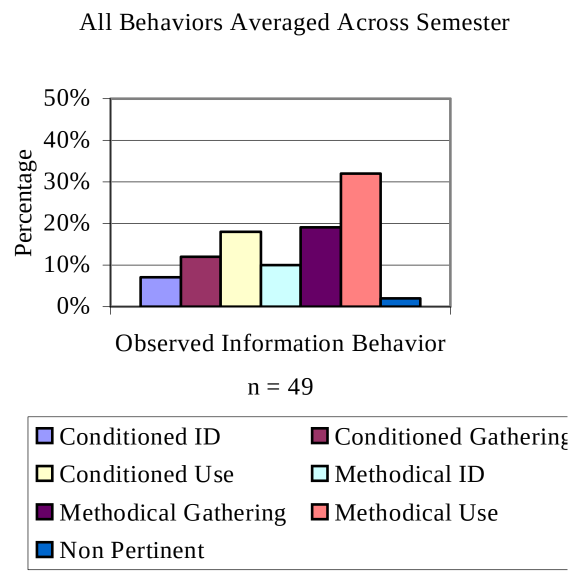


Figure 2. *Percent total observed Crew Information Behaviors across semester.*

Using the performance criteria, a total of 16 missions (33%) were classified as high performance, 14 (28.5%) as average performance and 19 (38.5%) as low performance (Table 1).

Table 1. *Distribution of performance groupings by scenario for 49 missions.*

Mission	Team Performance			
	High	Avg	Low	Total
Scenario 1	5	2	1	8
Scenario 2	3	6	3	12
Scenario 3	3	3	6	12
Scenario 4	1	1	2	4
Scenario 5	4	2	7	13
Total	16	14	19	49

Among the high performance grouping, MU accounts for 36% of the behavior, MG is 21%, MI is 10%, CU is 17%, CG is 8%, CI is 5%, and NP is 3%. Non-pertinent and methodical use behaviors were displayed more frequently in the high performance missions than in any other. These missions also contained the lowest frequency of observed behaviors for conditioned identification, conditioned gathering, conditioned use and methodical identification (see Figure 3). Among the average performance grouping, MU accounts for 32% of the behavior, MG is 19%, MI is 11%, CU is 17%, CG is 12%, CI is 8%, and NP is 1%. Methodical identification, and methodical gathering behaviors were displayed more frequently in the average performance missions than any other. Among the low performance grouping, MU accounts for 28% of the behavior, MG is 16%, MI is 9%, CU is 21%, CG is 16%, CI is 9%, and NP is 1%. Conditioned identification, conditioned gathering, and conditioned use behaviors were displayed more frequently in the low performance missions than any other missions.

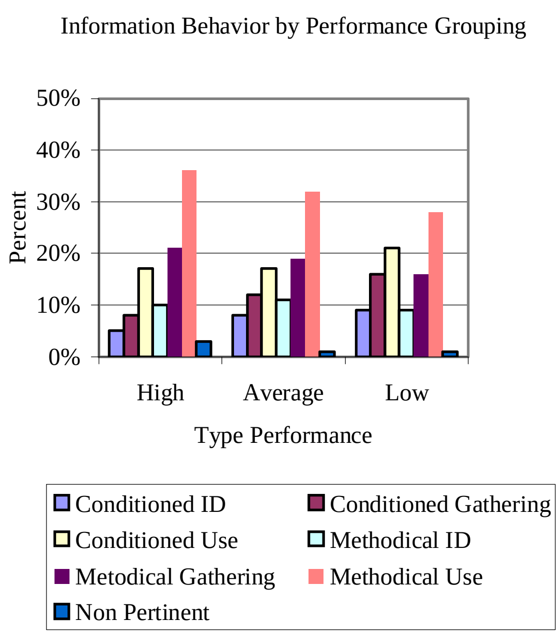


Figure 3. *Percent information behavior by performance grouping for 49 missions.*

The average number of information behaviors was smallest for the high performance missions, and largest for the average performance missions. From the data it appears high performance missions have less frequent information interaction overall (*M* = 229.44), followed by low performance missions with more frequent information interaction (*M* = 242.63), and average performance with the highest overall information interaction per mission (*M* = 256.29). This is not surprising since a tight coupling of activities between crewmembers, allowing a comprehensive shared understanding, marks high performance. Average and low performing teams appear not to share this tight coupling of activity or cohesive representation of the environment, resulting in the need for more interaction at lower fidelity.

A total of 8 missions resulted in an accident. All accident missions are contained within the low performance crew grouping, yet all low-performance crews did not have an accident. A means comparison between the three performance groupings reveals flight hours may account for significant differences between high and low performing crews, but does not appear as a significant difference between low and average crews or average and high crews (Table 2).

Table 2. *Comparison of crew flight hours between performance groupings ($p<.05$).*

Flight Hours						
Crew Performance	N	Mean	Std Dev	t	df	Sig.
High	14	496.68	255.55			
Average	12	345.88	187.81	1.69	24	.104
High	14	496.68	255.55			
Low	17	312.65	71.80	2.61	14.70	.02
Average	12	345.88	187.81			
Low	17	312.65	71.80	.584	13.29	.569

Analyzing the missions for crew performance factors in addition to information behavior results in a chi-square distribution revealing statistical significance (X^2_{12} = 320.62, $p<.001$). In particular it can be assumed that crew information behavior differs in relationship to crew performance.

Conclusion

When viewed as a whole, the observed information behaviors display a pattern in which methodical information behaviors are higher than that of their conditioned counterparts, with a low amount of extraneous non-pertinent chatter. It is reasonable to expect higher instances of methodical information behavior in aviation operations during the approach segment of the flight due to the prevalence of procedures and checklists. Since this portion of the flight also represents a period of higher workload, whether in the presence or absence of a system malfunction, crews require clearly defined processes and information that is easily accessible (Sarter & Woods, 1991). Information provided by and obtained through the use of checklists and procedures would necessarily be conspicuous as methodical information behaviors. Crews who properly employ checklists and procedures will more than likely have a higher incidence of methodical behavior, exploiting information processes. Crews who are more casual about procedures or who are not so comfortable with the airplane instrumentation may employ more conditioned "seat-of-the-pants" information behaviors and explore more avenues of potential information rather than exploit formal processes. The most significant differences though, lie between the way high performing crews act as a team in the negotiation of information meaning, and the way low performing crews (successful and accident) contend with information meaning. There are higher amounts of conditioned behaviors and lower amounts of methodical information behaviors among low performing groups than high performing groups.

What the proper proportion of information activity may be has yet to be determined through continued research. It appears overly methodical, or information exploiting, behaviors to the detriment of conditioned, or information exploring behaviors, may lead a crew to overlook the discovery new information that may contain cues that their previous assessment of the situation was flawed. However, the reverse appears to hold true also. Overly casual, conditioned information exploration behavior may lead a crew to mis-perform critical action sequences necessary for flight safety (see von Thaden, 2004). The balance remains to be determined, but this research approach may lead to the demonstration of the equilibrium in the information practice of high performing crews.

This study has shown it is possible to discern differences in the information practice of accident and non-accident involved flight crews. Effective information practice involves engaging in a variety of information behaviors that span across the 6 categories of the IBG. The IBG is a useful tool for understanding distributed information practice among flight crews, which may in turn inform improved crew resource management training and accident investigation. This framework also has the portability to be applied in other high risk, safety critical domains where work is performed through distributed collective practice, such as healthcare, nuclear power, and space exploration.

References

Buckland, M. (1991). Information retrieval of more than text. *Journal of the American Society for Information Science*, 42: 586-588.

Choo, C.W., Detlor, B., & Turnbull, D. (2000). Information seeking on the web: An integrated model of browsing and searching. *First Monday*. Issue 5.

Ellis, D. (1989). A behavioral model for information retrieval system design, *Journal of Information Science*, 13 (4 & 5).

Ellis, D., Cox, D., & Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of Documentation*, 49(4).

Ellis, D., & Haugan, M. (1997). Modeling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of Documentation*. 53 (4), pp. 384-403.

Hutchins, E. (1996). *Cognition in the wild*. Cambridge, MA: MIT Press.

Marchionini, G. (1995). Information seeking in electronic Environments. *Cambridge Series on HCI*. Vol 9. Cambridge: Cambridge University Press.

Sarter, N.B., & Woods, D.D. (1991). Situation awareness: A critical but ill-defined phenomenon. *International Journal f Aviation Psychology*, (1) p 45-57. New Jersey, LEA.

Weick, K., & Roberts, K. (1993). Collective mind in organizations: Heedful interrelating on the flightdeck. *Administrative Science Quarterly*. 38(3), pp. 357-381.

Wilson, T.D. (2000). Human information behavior. *Informing Science*. 3 (2) pp. 49-56.

von Thaden, T.L. (2003). Information behavior among commercial aviation CFIT accident flightcrews: Transcript analyses. *Proceedings of the 12th International Symposium on Aviation Psychology*. Dayton, OH, April 14-17.

von Thaden, T.L. (2004). Developing a methodology to study crew information behavior in aviation. Technical Report AHFD-04-13, University of Illinois, Aviation Human Factors Division.

von Thaden, T. & Wiegmann, D. (2004). Information behavior among accident-involved commercial aviation flight crews. *Proceedings of the Safety Across High-Consequence Industries Conference*, St. Louis, Missouri. March 9 & 10.

ORGANIZATIONAL FACTORS IN COMMERCIAL AVIATION ACCIDENTS 1990-2000

Terry L. von Thaden and **Douglas A. Wiegmann**
University of Illinois at Urbana-Champaign
Champaign, IL

Scott A. Shappell
Federal Aviation Administration Civil Aeromedical Institute
Oklahoma City, OK

Recently several major transportation accidents have brought significant attention to the role of organizational factors in supporting safety within high-risk critical systems. However, little is essentially known about the types of organizational factors that contribute to these accidents, as there has yet to be a comprehensive analysis of these factors. This paper elaborates on the types of organizational factors that have contributed to pilot-error related aviation accidents in U.S. commercial aviation. Specifically, we analyzed 60 accidents with organizational cause factors from 1990-2000. Results from this analysis indicate that the type and frequency of organizational factors that contribute to accidents varies across type and size of aviation operations. However, the data also argue for a more thorough analysis of organizational factors during an investigation so that a clearer understanding of the actual contributing factors to an accident involving pilot error can be discerned.

Introduction

Organizational factors play a significant role in the foundation of safety in high-risk systems. Several high profile accidents in the late twentieth century brought considerable attention to the role of organizational factors in accident causation. One of the first instances was the nuclear accident at Chernobyl in 1986. The International Atomic Energy Agency identified a “poor safety culture” as a factor contributing to this disaster (IAEA, 1986, as cited in Cox & Flin, 1998; Pidgeon, 1998). Since that time, organizational factors have been discussed in other major accident enquiries and analyses of system failures such as the King’s Cross underground subway fire in London and the Piper Alpha oil platform explosion in the North Sea (Cox & Flin, 1998; Pidgeon, 1998). Organizational factors also began to appear in the discussions of several high profile aviation/aerospace accidents such as the Challenger disaster (Vaughan, 1996).

The turning point for the analysis of organizational factors within commercial aviation accidents came with the National Transportation Safety Board’s (NTSB) report of the in-flight structural breakup and crash of Continental Express Flight 2574 near Eagle Lake, Texas, on September 11, 1991 (Meshkati, 1997). One Board member, in a dissenting opinion, suggested that the probable cause of this accident included, “The failure of Continental Express management to establish a corporate culture which encouraged and enforced adherence to approved maintenance and quality assurance procedures” (NTSB, 1992, p.54). Since then, the focus on organizational factors in aviation and aerospace accidents has continued to expand to include the recent analysis of the organizational failures within NASA that contributed to the Columbia Space Shuttle tragedy (CAIB, 2003).

The role organizational factors play in the etiology of accidents has been acknowledged prior to the aforementioned accidents. For example, March and Simon (1958), in their influential work *Organizations*, describe organizations as complex systems whose failings are more often directed at administrative factors, rather than at operator (worker) behavior. Likewise, Heinrich, Peterson, and Roos (1959), discuss organizational opportunities for accident prevention efforts in their work *Industrial Accident Prevention*. Bird’s (1974) Domino Theory fundamentally traces the root causes of all accidents to failures in organizational loss control and has been a standard model of accident causation within industrial and manufacturing settings for decades. More recent theories of organizational accidents build on these and other foundations, including works by Reason (1990; 1997), Weick and Roberts (1993), Klein, Bigely, and Roberts (1995), and Zhuravlyov (1997).

But what is actually known about the types of organizational factors that contribute directly to accidents, namely commercial aviation “pilot error” type accidents? There is a growing body of knowledge in relation to the role that aircrew or pilot error plays in the cause of aviation accidents. For years, the unsafe actions on the part of the pilots as accident causal factors have hovered around 80% (Dismukes, Young, & Sumwalt, 1999; Wiegmann & Shappell, 2003). This is not surprising since pilots’ actions are more easily tied to the occurrence of an accident, whereas organizational factors are generally far removed in time and space from an accident, making them difficult to link to an accident during an investigation (Wiegmann & Shappell, 2001). In addition, accident investigators are often highly knowledgeable of the tasks and duties of the accident aircrew that may have gone awry, but may be generally uninformed as to the types of organizational issues that they should specifically examine during an investigation.

Accordingly, there is debate that despite a growing awareness of the importance of organizational factors, they have been often overlooked or unidentified by aviation accident investigators in the field (Yacavone, 1993; Maurino, Reason, Johnston, & Lee, 1995). That is, most field investigations refer to the pilot's erroneous decision or action with little understanding of the contributing factors committed by those within the organizational chain of command. This is not to say that aviation accidents may be completely devoid of causal factors on the part of the pilot(s), but rather to note that the emphasis most often has been placed on the frontline operators, rather than tracing back up the organizational chain.

This paper elaborates on the types of organizational factors that have contributed to commercial aviation accidents in the U.S. Specifically, we analyzed 60 accidents with organizational cause factors from 1990-2000.

Method

We analyzed the NTSB's commercial aviation accident data for the ten-year period from 1990-2000. This set of accidents includes Federal Aviation Regulation (FAR) Part 121 scheduled and non-scheduled operations and FAR Part 135 scheduled and non-scheduled operations. From the original set of 1322 commercial aviation accidents, 781 were identified as having human factors causes using the Human Factors Analysis and Classification System (HFACS) (Wiegmann & Shappell, 2003). Sixty of these accidents attributable at least in part to pilot error contained 70 organizational cause factors. A comprehensive analysis of these accidents was performed using the NTSB's assigned findings. (Note: accidents relating to organizational factors associated with maintenance facilities and maintenance issues were not included in this analysis.)

Results

Of the 60 identified accidents, 73% produced some type of injury or fatality, while only 27% resulted in no injuries to crew or passengers (Table 1).

Table 1. *Degree of injury sustained from 60 commercial aviation accidents with organizational cause factors, 1990-2000.*

	Frequency	**Percent**
None	16	27%
Minor	8	13%
Serious	7	12%
Fatal	29	48%
Total	60	100%

Within these accidents, 17 occurred in FAR Part 121 aviation operations, while 43 occurred in FAR Part 135 aviation operations. When broken down into type of hauling operation these accidents represent under each certificate of operation, passenger-only operations make up the largest category of accidents, followed by cargo-only operations and then passenger-cargo combined operations. A comparison of the type of hauling operation these accidents represent under each certificate of operation is presented in Table 2.

Table 2. *Comparison of type of hauling operation in 60 organizational accidents, 1990-2000.*

		Passenger Only	**Cargo Only**	**Passenger/ Cargo**
Part 121	Scheduled	7	2	5
	Non-scheduled		3	
Part 135	Scheduled	8		2
	Non-scheduled	21	11	1

Assessing Organizational Factors

Assessing the assigned findings for the accident sequence of events provided a more complete analysis of the 70 organizational factors associated with the 60 accidents. We used the NTSB identified accident sequence of events identified during the original investigation. Based both on the descriptors provided by the NTSB and a review of the narratives associated with each of the factors, we were able to cluster these organizational factors around 10 broad categories which include procedures, training, surveillance, standards, information, supervision, pressure, documentation, substantiation, and facilities. A brief description of these factors appears in Table 3.

Table 3. *Organizational contributing factors of 60 U.S. commercial aviation accidents 1990-2000 (clustered).*

Category	Description
Inadequate procedures or directives	Ill-defined or conflicting policies Formal oversight of operation
Inadequate initial, upgrade, or emergency training/transition	Opportunities for pilot training not implemented or made available to pilots (e.g., human resource problem)
Inadequate surveillance of operations	Organizational climate issues Chain-of-command Quality assurance and trend information
Insufficient standards/requirements	Clearly defined organizational objectives Adherence to policy
Inadequate information sharing (untimely or insufficient)	Logbooks, updates, and weather reports on the part of the organization
Inadequate supervision of operations (management level)	Failure to provide guidance, oversight, and leadership to flight operations
Company/management induced pressure	Threats to pilot job status and/or pay
Faulty documentation	Inaccurate checklists, signoffs, and company record keeping that effects flight operations
Inadequate substantiation process	Well-defined, verified process Accountability Standards of operation Regulation Recording/reporting process
Inadequate facilities	Failure to provide adequate environmental controls, lighting, clearance, etc. for flight operations

When these organizational cause factors are considered in relation to operational category (Table 4), a clearer picture of the elements related to aviation operations emerges.

Accident factors related to inadequate organizational procedures emerge prominently in both Part 121 and Part 135 operations, with 7 instances (9.5%) in Part 121 and 8 instances (11.5%) in Part 135 operations. The factors associated with inadequate training are significantly higher in Part 135 operations (16%), than in Part 121 operations (3%). Inadequate surveillance of operations also ranks higher in Part 135 operations (10.5%) than in Part 121 operations (3%), as do inadequate standards/requirements at 9% and 3%, respectively. Inadequate information sharing ranks higher in Part 121 accidents (7%), than in Part 135 operations (4.5%). Accident factors associated with inadequate supervision, which includes management oversight, are present in Part 135 operations (10.5%) but not in Part 121 operations, as are factors associated with company-induced pressure (6%) and inadequate facilities (1.5%).

Discussion

A strong reason for the discrepancy of accident distribution between the operative categories could lie in the range of pilot non-flight duties, which depends on the employment setting. Part 121 airline pilots have the services of large support staffs, and consequently perform few non-flight duties. Pilots employed in other settings, such as Part 135 operations have duties other than flight responsibilities. They may load the aircraft, handle passenger baggage, supervise refueling, arrange for major maintenance, or perform minor aircraft maintenance and repair work.

This leads to a blurring of the supervisory chain of command and can put one person in charge of numerous supervisory issues, devoid of checks and

Table 4. *Cross-tabulated breakdown of 70 organizational contributing factors to 60 commercial aviation accidents 1990-2000.*

	Part 121 Scheduled	**Part 121 Non-scheduled**	**Part 135 Scheduled**	**Part 135 Non-scheduled**	**TOTAL**
Procedural	8% (6)	1.5% (1)	1.5% (1)	10% (7)	21%(15)
Training	3% (2)		12% (8)	4% (3)	18%(13)
Surveillance	1.5% (1)	1.5% (1)	1.5% (1)	9% (6)	13% (9)
Standards	3% (2)		3% (2)	6% (4)	12% (8)
Information	4% (3)	3% (2)	1.5% (1)	3% (2)	12% (8)
Supervision			1.5% (1)	9% (6)	10% (7)
Pressure				6% (4)	6% (4)
Documentation	3% (2)			1.5% (1)	4% (3)
Substantiation	1.5% (1)		1.5% (1)		3% (2)
Facilities				1.5% (1)	1.5% (1)

Percentages are approximate due to rounding.

balances, which they are not adequately equipped to handle. This may also serve as a contributing factor to the higher rate of inadequate supervisory and surveillance accident factors at the Part 135 operations than at the Part 121 operations.

As airlines grow larger, the problems appear to display tendencies shifting from those of direct supervisory and pressure, to those of a procedural, informational, documentary nature. What this may represent is a drift in the practical application of safety concepts. Normal rote operations may shift from time to time based on the accepted way work is performed. These shifts may also become part of organizational doctrine, as the safety rules for the original procedure become lost in the presence of the current context of work. This conceptual drift appears to contribute to the organizational factors experienced in the larger air carriers where procedural departures from routine become routine in practice in the absence of documentation and information sharing. This may be due to the hierarchical distance between the front line operators and the upper level management where the procedure is substantiated.

An abundance of factors occur toward the top of the organizational chain. Indeed, problems with the organization's procedures were cited in a majority of the accidents studied. The overarching organizational process set by those in charge of establishing the organization's directives and procedures may come into play that those in charge of setting policy are too far removed from the actual job to adequately address the issues involved. Perhaps it behooves those in charge, in the policy area specifically, to be sure a more bottom-up organizational approach is utilized to incorporate the expertise of those who actually perform the work with that of those who preside over it.

Conclusion

This research provides an overview of the concept of organizational safety as related to the human factors perspective. We introduce a framework to objectively identify organizational factors as related to pilot error accidents. Once organizational factors are identified, interventions aimed at the supervisory and organizational levels of an establishment have the potential to improve the entire system when compared to issues at the operator level, which may focus on alleviating just one error. Valuable resources are better spent on prevention and control at the organizational level, rather than on trying to fix, after-the-fact, the inexhaustible ways people fail at the operational level. With this, we have the potential to eliminate a myriad of errors as opposed to the proverbial Dutch boy putting his finger in the dam, only to find numerous leaks exploding all around.

It bears mention that the accidents presented here are assessed according to the NTSB's findings of probable cause. Other accidents may meet the criteria of containing organizational cause factors, yet organizational factors in accident investigations have been historically overlooked and thusly not directly traceable as such in any findings. As a result, we have not included them here, thus the number of organizational accidents in commercial aviation may be higher than reported here.

Acknowledgement

This work was supported in part by a grant from the Federal Aviation Administration under Award No.

DTFA 01-G-015. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the Federal Aviation Administration.

References

Bird, F. (1974). *Management guide to loss control.* Atlanta, GA: Institute Press.

Columbia Accident Investigation Board (2003). Report from the space shuttle Columbia accident investigation, Report Volume 1 downloaded from: http://anon.nasa-global.speedera.net/anon.nasa-global/CAIB/CAIB_lowres_full.pdf

Cox, S., & Flin, R. (1998). Safety culture: Philosopher's stone or man of straw? *Work & Stress*, *12*(3), 189-201.

Dismukes, K., Young, G., & Sumwalt, R. (1999). Cockpit interruptions and distractions: Effective management requires a careful balancing act. *Airline Pilot*, 68 (5).

Heinrich, H., Petersen, D., & Roos, N. (1959). *Industrial accident prevention: A safety management approach* (5th edition). New York: McGraw Hill.

Klein, R. L., Bigley, G. A., & Roberts, K. H. (1995). Organizational culture in high reliability organizations: An extension. *Human relations*, 48, 771-793.

March, J., & Simon, H. (1958). *Organizations.* New York: Wiley.

Maurino, D.E., Reason, J., Johnston, N., & Lee, R. (Eds.) (1995). *Beyond aviation human factors, safety in high technology systems.* United Kingdom: Ashgate.

Meshkati, N. (1997, April). *Human performance, organizational factors and safety culture.* Paper presented on National Summit by NTSB on transportation safety, Washington, D.C.

National Transportation Safety Board. (1992). *Aircraft Accident Report. Britt Airways, Inc., d/b/a Continental Express Flight 2574, In-flight Structural Breakup, EMB-12ORT, N33701, Eagle Lake, Texas, September 11,199l* (Rep. No. NTSB/AAR-92/04). Washington DC: Author.

Pidgeon, N. (1998). Safety culture: Key theoretical issues. *Work & Stress, 12*(3), 202-216.

Reason, J. (1990). *Human error.* New York, NY: Cambridge University Press.

Reason, J. (1997). Managing the risks of organizational error. Brookfield, VT: Ashgate.

Vaughan, D. (1996). *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. IL: University of Chicago Press.

Weick, K. E. & Roberts, K. H. (1993). "Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly*, 38, 357-381.

Wiegmann, D. A., & Shappell, S. A. (2001). Human error analysis of commercial aviation accidents: Application of the human factors analysis and classification system (HFACS). *Aviation Space and Environmental Medicine*, *72*(11), 1006-101.

Wiegmann, D. A., & Shappell, S. A. (2003). *A human error approach to aviation accident analysis: The human factors analysis and classification system.* VT: Ashgate.

Yacavone, D. (1993). Mishap trends and cause factors in naval aviation: A review of Naval Safety Center data, 1986-90. *Aviation, Space and Environmental Medicine*, 64, 392-395.

Zhuravlyov, G. (1997). Social-Technical systems as a framework for safety culture. In *From Theory to Practice: Psychological Foundations of Safety Culture in Nuclear Installations.* Moscow: Central Economics and Mathematics Institute, Russian Academy of Sciences.

THE NEED FOR QUALITY AVIATION SAFETY GRADUATES: AN EDUCATIONAL CHALLENGE FOR THE 21ST CENTURY

Thomas R. Weitzel
Embry-Riddle Aeronautical University
Daytona Beach, Florida

Thomas R. Chidester
NASA Ames Research Center
Moffett Field, California

The authors investigated a previously unaddressed problem within the curricula of the United States (U.S.) aviation institutions of higher education. Graduates of these institutions were not being prepared to work within the safety departments of the U.S. air carriers involved with one or more of the five current, voluntary programs. To ascertain the need for a solution, a subjective instrument was developed and personally administered to 13 participants within the industry. The qualitative results were interpreted, and, in combination with the knowledge gained from the immersion of a professor within a research organization, resulted in placement of some of the aforementioned content within the curriculum on one campus of one U.S. aviation university in the spring of 2005.

Introduction

The purpose of this paper is to serve as a progress report for the project that was conceived by the authors and enabled by the National Aeronautics and Space Administration (NASA) Faculty Fellowship Program. During the summer of 2004, a professor (also a retired air carrier captain) from Embry-Riddle Aeronautical University (ERAU) and a NASA Ames Research Center program manager commenced a project that would infuse the five, current Federal Aviation Administration (FAA) and U.S. air carrier voluntary safety programs into curricula for aviation institutions of higher learning. Current Air Traffic Control (ATC) programs that address improved efficiency of the National Airspace System (NAS) have not been excluded from the program. They have been scheduled for further investigation during the summer of 2005 as content for a proposed ATC specialization in the Master of Science in Aeronautics (MSA) program at ERAU, Daytona Beach, Florida.

The safety programs are relatively new to the U.S. air transportation system, having matured only since the 1990s. During the 21st century, the FAA and NASA have emphasized a need for the continuous, reliable analysis of the program-derived, large safety/ efficiency databases of both the U.S. air carrier and ATC systems. Two generalizations concerning the analysis, interpretation, and reporting processes associated with the large volumes of data are:

1. The air carrier personnel traditionally involved with the analysis and reporting of the data generated by today's modern safety programs possess considerable operating experience, but have not had scientific backgrounds.
2. The U.S. aviation institutions of higher learning have not had the resources to introduce the new air carrier safety programs to the curricula. More explicitly, the knowledge of, and the materials for, the programs have not been available.

Thus, it was theorized that the NAS would benefit from future graduates of aviation higher education with the desirable scientific knowledge, skills, and attributes associated with the maturing safety and efficiency programs of the 21st century.

Background

Less than 5 years after the Wright Brothers' first controlled, powered flight, the U.S. experienced its first aviation passenger fatality. On September 17, 1908, Orville Wright was demonstrating the Wright flying machine to U.S. Army officials, with a passenger. The aircraft crashed, with resultant fatal injuries to the passenger, an Army Lieutenant (Thomas Etholen Selfridge, n.d.). More public scrutiny occurred when Knute Rockne, a popular football coach from the University of Notre Dame, was killed in a 1931 accident, followed by the 1935 fatal aircraft accident that killed U.S. Senator Bronson M. Cutting of New Mexico (Komons, 1989). Prior to the 1938 Civil Aeronautics Act that derived from these two 1930s accidents involving notorieties, ". . . the mantra seemed to be 'fly it, crash it, redesign it, fly it, crash it . . .' resulting in only modest improvements over time" (Walters, 2002, p. 2). From 1938 through 1974, the U.S. regulation of aviation and the investigation of accidents became structured such that: (a) the FAA is housed within the U.S. Department of Transportation (DOT), and (b) an agency separate and independent of the DOT, the National Transportation Safety Board (NTSB), is assigned the investigation of serious incidents and accidents. A positive result was that the charting of the air carrier accident rate became asymptotic. Then, in 1990, a representative of The Boeing Company announced "If the current rate stays absolutely flat, a projection based on the increase in the number of airplanes in service shows that, by the year 2005, there will be an airline hull loss somewhere in the world approximately every two weeks" (Weener, p. 1). It was an understatement to classify this projected statistic as 'unacceptable to the public.'

Preparing for the 21st Century

Due to very low frequencies, analysis and reporting of accidents and serious incidents have not been good metrics of the NAS system safety. In 2003, the FAA reported that the probability of an air carrier accident per departure/flight was less than $.3 \times 10^{-6}$. Current media reports quote the FAA and NTSB as stating that the rate for 2004 was $.15 \times 10^{-6}$ (Miller, 2005). The industry has recognized the need to look for precursors of accidents in events detectable in routinely-recorded data, reported by operational personnel, observable in training performance, or in disciplined audit of airline safety processes.

In 1975, the FAA and NASA signed a Memorandum of Agreement that established the Aviation Safety Reporting System (ASRS), with NASA responsible for the design and implementation of the incident-reporting program. The ASRS has collected, analyzed, and responded to voluntarily submitted aviation safety incident reports in order to lessen the likelihood of all aviation accidents. This has been particularly important as the literature has generally conceded that over two-thirds of all aviation accidents and incidents have their roots in human performance errors.

In the 1990s, the air carrier industry joined with the FAA and NASA in addressing the problem of further decreasing the airline accident rate as the volume of air traffic grew. Collaborating with innovative airline initiatives, the FAA introduced five air carrier safety partnership programs, which are administered by its AFS-230 office and are maturing in the 21st century. The goal of each program is continued improvement for an already very safe U.S. air transportation system. The five voluntary partnership programs, designed to be inter-related, are:

1. The Flight Operational Quality Assurance (FOQA) Program – de-identified digital data obtained from a Quick Access Recorder (QAR) are utilized to target and resolve safety issues.
2. The Aviation Safety Action Program (ASAP) – de-identified, employee self-disclosures are utilized to target and resolve safety issues.
3. The Advanced Qualification Program (AQP) – a training program that contains self-correcting quality assurance components and utilizes de-identified individual performance data to target and resolve training/safety issues.
4. The Internal Evaluation Program (IEP) – entails internal safety audits, in combination with documented organizational responsibilities, safety information acquisition procedures, and continuous quality assurance processes that are designed to increase the likelihood that safety deficiencies are promptly identified and corrected.
5. The Voluntary Disclosure Reporting Program (VDRP) – allows for corporate self-disclosure in identifying and resolving safety issues.

Together, the five programs have continued to generate both objective and subjective volumes of data, all of which require comprehensive analysis and interpretation before reporting. The air carrier FOQA programs have required quantitative data analysis, and several vendors have developed sophisticated data downloading and analysis programs. Subsequent to data validation, the statistical programs allow the creation of a database, to which statistical treatments can be applied. The treatments enable summarization, and interpretation of the data; data reduction within large databases has necessitated the implementation of multivariate statistical techniques. The air carrier AQP programs have required the treatment of both quantitative and qualitative data. The quantitative data have been analyzed with an appropriate statistical program (e.g. general statistical analysis packages, such as SPSS). The qualitative data have been treated in several manners; the implementation of a relational database (e.g., MS Access) appears to have been most appropriate.

The air carrier ASAP programs, with large volumes of subjective data, have generally led to analyses that required reduction, display, and verification of the data before arriving at any interpretations. Examples of the relational database programs that have been used are MS Access and Oracle. Detection of the relationships and hidden patterns in the subjective narratives has resulted in the implementation of text data mining programs (e.g., Clementine from SPSS and PolyAnalyst from Megaputer Intelligence). The IEP and the VDRP require self-auditing. Extensive records are involved with the IEP, and these involve both quantitative and qualitative data. The VDRP qualitative data are derived from self-audit tracking.

In cooperation with the FAA, NASA initiated the Aviation Safety Program (later modified to the Aviation Safety and Security Program [AvSSP]) as an outgrowth of the 1997 White House Commission on Aviation Safety and Security Report – "The Gore Report." The goal for AvSSP, tracked as part of the August 2000 FAA-NASA Integrated Safety Research Plan, is the development of tools that will reduce the fatal accident rate 80% by 2007, and 90% by 2022. Some of the hierarchically structured components of the AvSSP are the:

1. Project: System Safety Technologies.
2. Subproject: Aviation System Monitoring and Modeling (ASMM).
3. Flight Data: Aviation Performance Measuring System (APMS).
4. Radar Data: Performance Data Analysis and Reporting System (PDARS).
5.

APMS has been developing the next generation of tools used by air carrier FOQA personnel for flight data analysis and interpretation. PDARS has been developing networking and analysis tools used by ATC facility-level managers for radar data. The APMS and PDARS tools have analyzed and interpreted the normal, routine operations for situations and trends that might be precursors of incidents and accidents.

An Exploratory Approach

To address the need for curricular change at ERAU and introduce the voluntary safety programs to U.S. academic institutions, the first author was immersed within the APMS group at the NASA Ames Research Center, and some industry safety practitioners. An interview protocol with 11 safety program managers at 3 U.S. airlines with whom NASA had Space Act Agreements (SAA), and 2 individuals at a software vendor with a SAA, was conducted during a 2-week period in June 2004. The emphasis of the interview instrument was upon defining the needed skills for future employees working in air carrier safety; it was designed so as to be two pages in length and to result in a semistructured administration. The environments were familiar and comfortable for the interviewee (and the interviewer); both parties had the 2-page instrument in front of them, and were free to make any notes; and the order of discussion of the items was introduced as not being important. Generally, the interviewer moved back and forth between the first page (the personal data of the interviewee) and the second page (the air carrier's current practices) with his note-taking, while the interviewee occasionally glanced at the items as they were discussed. The personal data began with date, time, and name of the interviewee, and progressed through the corporate relationships with other stakeholders and the levels of control and decision-making for the interviewee.
The second page was titled "Current Carrier Safety/ Quality Practices" and comprised 10 items. One item addressed the "desired personal attributes" for the job of the interviewee; another addressed the "desired personal skills." Both of these items included short lists – attributes and skills, respectively. None of the desirable attributes and skills listed was disagreeable to the interviewees; several additions to the short lists were made by some of the interviewees.

The time for each interview was forecast to be 15-30 minutes; however, most made more time and enjoyed the discussion (average time with each interviewee approximated 45 minutes). The 'pencil-and-paper' notes were later entered into a word-processor at the earliest opportunity. Analysis of the resulting documents was done by hand, and consisted of tallying the responses to those items that directly related to the future students of aviation safety education. Investigation continued for an additional 8 weeks utilizing phone conversations and e-mail (and one data analysis working group meeting) with the air carrier personnel, software vendors, hardware manufacturers, and the FAA's AFS-230 office. Non-proprietary materials for course content were provided willingly by several of the individuals.

The sums of the replies to the qualitative queries confirmed an industry need for future safety employees versed in data acquisition, analysis, interpretation, and reporting required by the current safety programs. Curricular placement of the content was reasoned to be a course at the graduate level. Interviewees, and subsequent contacts, were in unanimous agreement that statistical knowledge and presentation skills were highly desirable, and that a course in a master's degree program appeared to best fulfill the requirements. Knowledge of the air carrier/aviation system and its components (e.g., ATC, operations, maintenance, and dispatch) would be a must. Communication skills, both oral and written, honed in a graduate program, were deemed beneficial. Unanimous agreement existed as to the desirable attributes (and abilities) of the students. The requisite credibility (mentioned by numerous interviewees and subsequent contacts) would demand trustworthiness, honesty, reliability, integrity, assertiveness, etc. In addition to familiarity with a number of software (ideally statistical and database) applications, several interviewees and contacts expressed the need for an understanding of computer logic. Skill with at least one programming language would be helpful. The proposed future integration and pooling of data from different software and servers reinforced the need for some knowledge of computer logic.

The ERAU Seminar

During the spring of 2005, an advanced graduate research course, utilizing the results of the summer 2004 fellowship, has been implemented as a graduate seminar in the MSA. The facilitators for the seminar have been the aforementioned researcher/professor and a U.S. major air carrier pilot doing his dissertation research for a doctoral program in adult education at another university. The course was capped at 12 students (11 actually enrolled), with the current core research and statistics course as the prerequisite (a course taught by the same professor). The five voluntary air carrier safety programs (including their interrelationships) serve as the archetype for the advanced research.

An appropriate text for this aviation-specific research seminar appeared to be the 'guide,' centered upon applied aviation research methods, by Wiggins and Stevens (1999). The research students have been assigned précis of the text's chapters, consisting of presentations as well as papers. Two chapters of the text provide a review of the statistical procedures (through the Analysis of Variance [ANOVA] and Chi-Square tests) that were course material in the prerequisite core research course. Power analysis and Principal Components Analysis (PCA) will be introduced with chapters from the multivariate text by Tabachnick and Fidell (2001). The process of data text mining has been addressed within student research assignments. Quasi-experimental research methods will be introduced with chapters from the 2002 design text by Shadish, Cook, and Campbell.

Multivariate analysis, specifically PCA, has been a statistical tool used by NASA's APMS for the FOQA program. Similarly, the PCA and ANOVA have been utilized in combination with survey and correlational techniques in addressing pilot safety and training (Baker, Beaubien, & Mulqueen, 2002; Hunter, 2005). The Baker et al. report also addresses the critical importance of qualitative analysis for those safety programs that provide subjective data (i.e., all but those generated by the QAR and FOQA).

The aforementioned, adult education Ph.D. candidate has obtained his committee's approval to continue as a seminar researcher/developer/instructor (with some attendant, self-developed evaluations of the seminar students) during the three semesters that the course has been scheduled to be offered as a developmental seminar. At the beginning of this spring's first seminar, a pretest of knowledge in several domains that would be desirable for future safety personnel in the air carrier industry was developed and validated by three researchers/practitioners from industry. The pretest was administered to the 11 master's students (mean ages and years of aviation experience were 28.91 and 8.32, respectively) during the first hour of the seminar's first meeting on January 14, 2005. The posttest, utilizing the same instrument, is scheduled to be administered during the first hour of the last meeting on April 22, 2005. The same pretest-posttest instrument will be used during the fall 2005 and spring 2006 semesters. Limiting confounds do exist with this design (Campbell & Stanley, 1963). Some of these are maturation, pretest sensitization, and differential selection (although the course is not currently required of any students). A history effect – the measurement is being performed three times with three groups over a timeframe of 16 months – in combination with the aforementioned threats and the pretest-treatment interaction weaken the validity of this pre-experimental design. In spite of a less than robust study, we believe that the data will reveal a favorable trend and lend support to the theory that academia can be of assistance in the preparation of future air carrier quality safety personnel.

Following the pretest, the first seminar meeting featured a discussion led by Dr. Douglas Farrow of the FAA's AFS-230. The relationships that exist between the programs were stressed in a manner that is currently nonexistent within the literature – a most valuable experience for the seminar. During the summer of 2004, there was no shortage of volunteers to speak in front of the graduate research seminar that would result from this project. Thus, other guests from the research community and industry have been scheduled to present before this spring's initial seminar. This 'access to expertise' has been designed to be a component of the students' research assignments. The students, in accordance with their interests, have been assigned to research the five programs. The resultant written reports, and presentations, will be compiled and distributed to the participants of the seminar on a Compact Disc (CD).

The 'computer logic and associated technology' that was mentioned by a number of industry's summer 2004 interviewees has been addressed by assigning a student (and manager of Information Technology on the campus) to research a course solution. (It appears that future seminars/classes would benefit from similar students.) A précis of a suitable chapter, combined with two iterations of the research progress and the final report, will begin to address the goal of familiarity with a sequential programming language. Progress toward the achievement of this problematic, lofty, and worthy goal should bear some rewards along the way.

The adult-structuring of the seminar has enabled collaborative learning, exposure to expertise and technology, and a mentoring relationship versus the apprenticeship model traditionally associated with graduate students and professors (Brookfield, 1988; Bye & Henley, 2003). The current seminar students should possess the required advanced technical skills for future safety data analysis and interpretation.

Future Outcomes

The spring of 2005 has the multi-year project on track. The overall academic program at ERAU's Daytona Beach campus has the graduate seminar continuing to be offered in the fall of 2005 and the spring of 2006 as it is developed for inclusion in the fall 2006 catalog as a second research course in the MSA core. Graduate interns that are selected from the program should be more valuable to more organizations than those currently being provided by ERAU's MSA for the air carrier safety departments.

Recommendations

Within two of the five current MSA specializations, safety systems and human factors, there appears to be the need for 'stand-alone' course content that would combine the maturing air carrier safety programs with quality management (Stolzer & Halford, 2004.) Farrow (personal communication, January 14, 2005) noted that a new model (and its associated acronym) has been discussed – the Safety Quality Management System (SQMS). It is recommended that in the fall of 2006, with the second core research course in place, the development of a SQMS seminar be investigated. An ATC specialization within the MSA that would utilize PDARS has been recommended and is being considered for evaluation during the summer of 2005.

Additionally, it is recommended that the results and outcomes of this research be shared with other institutions of aviation higher learning. To that end, current plans call for presenting the progress of the overall project to members of the following organizations: the International Society of Air Safety Investigators, the Human Factors and Ergonomics Society, and the University Aviation Association.

References

Baker, D. P., Beaubien, J. M., & Mulqueen, C. (2002, October 10). *Airline pilot training survey final report*. Washington, DC: American Institutes for Research.

Brookfield, S. (1988). Graduate adult education as a sociocultural product: A cross-cultural analysis of theory and practice in the United States and Great Britain. In S. Brookfield (Ed.), *Training educators of adults: The theory and practice of graduate adult education* (pp. 279-314). London: Routledge.

Bye, J., & Henley, I. (2003). Characteristics of adult learners and adult learning principles: Implications for aviation education. In I. M. A. Henley (Ed.), *Aviation education and training: Adult learning principles and teaching strategies* (pp. 30-62). Burlington, VT: Ashgate.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.

Federal Aviation Administration (FAA). (2003). *FY 2002 performance and accountability highlights.* Washington, DC: Author.

Hunter, D. R. (2005). Measurement of hazardous attitudes among pilots. *The International Journal of Aviation Psychology, 15,* 23-43.

Komons, N. A. (1989). *Bonfires to beacons: Federal civil aviation policy under the Air Commerce Act, 1926-1938*. Washington, DC: Smithsonian Institution.

Miller, L. (2005, January 4). Airlines have good safety record for 3d [*sic*] straight year. *Boston Globe*. Retrieved January 4, 2005, from http:// www.boston. com/news/nation/articles/2005/01/04/airlines_have_ good_safety_record_for_3d_straight_year?mode=PF

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Stolzer, A., & Halford, C. (2004). Wanted – quality in airline safety: Quality methods and tools are needed to manage new directions. *The Journal of Aviation/Aerospace Education & Research, 13*(3), 27-32.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn and Bacon.

Thomas Etholen Selfridge, First Lieutenant, United States Army. (n.d.). Retrieved July 2, 2004 from http://www.arlingtoncemetery.net/thomaset.htm

Walters, J. M. (2002). *The pilot proficiency audit of knowledge, skills, and abilities: Adding to the air carrier safety toolbox*. Unpublished master's thesis, Embry-Riddle Aeronautical University, Daytona Beach, FL.

Weener, E. F. (1990, December). *Control of crew-caused accidents*. Paper presented at Third Annual Society of Automotive Engineers Symposium on Human Error, Dallas, TX.

White House Commission on Aviation Safety and Security. (1997). *Final report to President Clinton*. Retrieved July 2, 2004, from http:// www. fas.org/irp/threat/212fin~1.html

Wiggins, M. W., & Stevens, C. (1999). *Aviation social science: Research methods in practice*. Brookfield, VT: Ashgate.

ATTENTIONAL TUNNELING AND TASK MANAGEMENT

Christopher D. Wickens
University of Illinois, Aviation Human Factors Division
Savoy, Illinois

This paper discusses attentional tunneling as one cause of breakdowns in task management. The phenomenon is defined, and empirical evidence is then reviewed to show the conditions in which the phenomenon is created by head up display location, compelling 3D displays, fault management, and automation induced complacency. Statistical and methodological issues are reviewed regarding the generalization of the phenomenon in the laboratory to real world mishaps.

Introduction

Breakdowns in task management and task prioritization have been well documented to cause mishaps in aviation (Funk, 1991; Chou, Madhavan, & Funk, 1996). A classic accident here is the crash of the Eastern Airlines L1011 into the Everglades, when pilots failed to manage their descending altitude while addressing an apparent landing gear failure. While such breakdowns have diverse psychological causes (Dismukes, 2001), our specific interest in this paper is focused on a collection of related phenomena that are known variously by names of "attentional tunneling", "attentional fixation" or "cognitive tunneling". Note that in this context, "attention" and "cognition" can be used nearly interchangeably, if it is assume that attention can be directed both inward to cognition, as well as outward toward particular channels and events in the environment.

We can offer a rough definition of attentional tunneling as the allocation of attention to a particular channel of information, diagnostic hypothesis or task goal, for a duration that is longer than optimal, given the expected cost of neglecting events on other channels, failing to consider other hypotheses, or failing to perform other tasks. Thus note that the definition must include both the forces that "lock the tunnel" to its current channel, as well as a definition of a channel of neglect.

Such a definition can account for more specific mishaps in a wide variety of circumstances. For example automobile accidents while on the cell phone can be attributed to undesirable "engagement" in the process of generating and understanding conversations (Strayer & Johnston, 2001; Horrey & Wickens, in press). Analysis of the Three Mile Island nuclear power accident associated the crisis with operators' excessive tunneling on one (incorrect) hypothesis as to the nature of the obvious failure, and this hypothesis led them to fail to attend to contraindicating visual cues. The Air Force has identified attentional tunneling as being a major cause of F16 mishaps, and indeed a case can be made that nearly all CFIT accidents (Shappell & Wickens, 2003) can be associated with attentional tunneling away from important altitude information.

While salient mishap data clearly indicates that the tunneling problem exists, such data often provide little usable evidence about its precise causes, because of the invariable absence of control that such data contain when they are used retrospectively to infer causality. Thus a complementary approach is to turn to more controlled flight simulation experimental data to both reveal the prevalence of the phenomenon in the general population, as well as the causal factors that amplify the likelihood of tunneling. Below we describe empirical data that bear on proposed causes of attentional tunneling to examine how the literature supports the degree of influence of each. We focus explicitly on four different factors that have been postulated to induce such tunneling: head up display location, the compellingness of 3D displays, fault management, and automation. We conclude with discussion of some of the methological and statistical issues involved in relating tunneling to flight safety.

Display location: HUD-induced tunneling. The now classic experiment of Fischer, Haines and Price (1980) revealed that pilots flying with a HUD were less likely to detect an unexpected runway incursion than those flying with conventional head down instruments, despite the fact that the HUD generally preserved the runway within foveal vision, where the incursion could be seen. While their observation of this phenomenon was not based upon a sufficiently large sample of pilots to reveal statistical trends, the phenomenon has been sufficiently replicated in both low fidelity (Wickens & Long, 1995) and high fidelity (Fadden, Ververs & Wickens, 2001; Hofer, Braune, Boucek, & Pfaff, 2000) simulations to establish it as real. Something about the HUD appears to attract attention to its image, and therefore lead attention away from important, but unexpected events within the visual field (see Wickens, Ververs

& Fadden, 2004 for a summary). Such HUD costs appear to be restricted to noticing totally unexpected events, since HUD benefits are generally found for most other visual tasks, including the detection of low frequency (but not truly surprising) events (Fadden et al., 2001).

3D Immersion Compellingness. The gradual appearance of 3D displays in the cockpit, such as the SVS guidance system (Prinzel et al., 2004; Schnell et al., 2004), has led to some concern that the highly realistic ego-referenced perspective of such a system can alter pilots' scan patterns so that they look extensively at the display (attentional tunneling), and fail to adequately sample the outside world. Such behavior can compromise safety to the extent that critical events, unknown to the data-generation sensors and software that drive the display, may be present as hazards in the outside world (e.g., the "rogue airplane" with an inoperable transponder; Wickens et al., 2002). Earlier research by Olmos Wickens and Chudy (2000) revealed such a trend exhibited in a 3D display in fighter aircraft in a low fidelity simulation. Four recent experiments in our laboratory described below, all using a high fidelity light aircraft Frasca simulator, clearly document the phenomenon.

Fadden, Ververs and Wickens (2001), compared a 3D "pathway-in-the-sky" display in a HUD location with a conventional HUD presenting ILS information in an approach and landing simulation. While we observed superior overall performance with the 3D display we did observe that the pathway induced a marginally significant 4 second delay in pilots' response to an unexpected runway incursion on a single (last) landing trial of the experiment.

Wickens, Alexander, Horrey, Nunes, and Hardy (2004; Thomas & Wickens, 2004), examined the guidance offered by a photo-realistic SVS display coupled with a 3D flight path pathway display in a long curved step down approach through a terrain challenged environment. Guidance and traffic detection performance with the 3D pathway was compared with that supported by less compelling (but equally accurate) instruments presenting the same flight path information. While flight path performance was much better supported by the integrated pathway, the detection of two unexpected or "off-normal" events was not. These included a blimp, located in the airspace on the flight path, but not visible in any head down display, and a runway offset, whereby the positioning of the SVS pathway and the synthetic runway on the display brought the pilots on an approach parallel to but offsetting the true runway (a disparity only detectable by looking outside).

We observed that 4 of the 8 pilots flying with the pathway failed to detect the blimp, whereas only 1 of 6 pilots flying without the pathway missed this critical off-normal event. Furthermore, while the runway offset was only imposed on those landing with the 3D pathway (and hence data could not be compared with those flying with the conventional instruments), 5 of the 12 pilots landing with the 3D pathway failed to detect the offset until very late in the landing phase. Furthermore, analysis of visual scanning revealed that the breakdowns in detection were associated with pilots who spent relatively more time looking head down at the instruments, rather than scanning outside. To some extent this head down scanning was "encouraged" by the rich and precise guidance offered by the pathway, and by the runway depiction on the head down terrain display lying on the SVS panel.

In a third study, Alexander, Wickens and Hardy (in press) also examined SVS-induced tunneling, although they did not compare their off-normal event detection with a control non SVS condition. On the final approach in their simulation, during the final trial of the experiment, a truly surprising runway incursion was present. This incursion did not itself form the basis of the unexpected event, since the tunnel guidance was designed to automatically reconfigure to form a missed approach path, and guide the pilot away from the runway obstacle. However the missed approach path was designed to put the flight trajectory squarely in the path of a blimp, visible only in the outside world, as in the first off-normal event examined by Wickens, Alexander, Horrey, Nunes, and Hardy (2004). Importantly, 17 of the 24 pilots in the experiment failed to detect the blimp, flying a flight path directly through it.

While the above findings suggest that the 3D pathway (and its associated SVS background) can inhibit the detection of truly surprising events, it is important to highlight two findings that failed to indicate "pathway induced tunneling". First, Wickens, Alexander, Horrey, Nunes, and Hardy (2004) examined a third off-normal event, a radio tower constructed so that it protruded into the pathway-defined flight path, but was visible on the SVS display. Here all pilots appeared to detect the tower adequately, as inferred from their flight path maneuvering.

The second example of "3D pathway success" was an experiment by Iani and Wickens (2004), using the same flight simulation as above, in which pilots' response to unexpected weather changes on a head down electronic weather map designed to influence the choice of an optimal safe flight path, were used to infer tunneling. Under these circumstances, those pilots flying with the 3D pathway display, which we hypothesized might induce tunneling, were actually more likely to notice the weather changes, than those flying with the separated instruments. This result, in seeming contradiction to the 3D pathway costs described above, were accounted for by two factors: (1) the weather changes, while unexpected, were not truly surprising, in that a well trained pilot, flying through areas where bad weather may exist, can be expected to be reasonably vigilant for unexpected changes in those weather patterns; (2) the 3D pathway was so much easier to fly (lower workload) than the separated display, that pilots were inferred to have a much greater amount of available attention with which to monitor the surrounding displays.

In summarizing these effects of immersed 3D display compellingness, we argue that some components of both a 3D SVS terrain background and a 3D pathway (or tunnel) hosted within, may contribute to a large allocation of visual attention to this location, an allocation which can leave a pilot vulnerable to missing truly surprising events that can only be seen elsewhere. Not all pilots demonstrate this, but those that do, tend to scan outside less than those that don't.

Importantly, one variable that appears to amplify this tunneling effect is the existence of a system failure. It is, for example, a failure of the overall SVS system that leads its guidance to a runway offset approach. Also, the one circumstance where the tunneling was most dramatically documented (over 70% of the pilots) was the finding of the blimp collision by Alexander et al. (in press), in what could be classified as a "double failure". That is, there was a runway incursion (failure of the air traffic management system), coupled with a failure of the SVS sensors to note the mid-air blimp following the missed approach path configuration. Thus we now discuss the contributions of failure management to attentional tunneling.

Failure Management. We noted above that attentional tunneling was amplified during the missed approach incident coupled with the sensor failure. Indeed there is a long history of research documenting the problems of failure and fault management inducing some sort of cognitive lockout, as true with the Eastern Airlines Everglades crash, and as demonstrated in other domains such as process control (Moray & Rotenberg, 1989). Dismukes (2001) has highlighted fault management as one of the "red flags" that pilots need to consider, as they remember to sample other non-fault-related instruments in the cockpit. The extent to which this results from the stress-induced cognitive narrowing brought on by the danger of the failure state (Hockey, 1986), or simply the high importance of the fault management task (which should optimally command a good deal of attention, even if not all of it) cannot be fully discriminated. Probably some of both factors are involved.

Automation Failure and Complacency. A final phenomenon, with great relevance to the cockpit, is that of automation induced "complacency" whereby a pilot, depending on automation which has always functioned safely in the past, fails to notice the unexpected failure (Parasuraman, Molloy & Singh, 1993; Parasuraman & Riley, 1997). This phenomenon is closely related to the "automation bias" reported by Mosier et al. (1998) whereby automation-based diagnosis is blindly followed by the pilot, in spite of evidence to the contrary. In a sense this phenomenon does not describe the capture of and "lock on" of attention (by the salient or compelling entity), so much as it describes the neglect of attention (to the channel characterizing the automated processing where events – failures -- are not expected to occur). Importantly, this phenomenon shares with other examples of tunneling described above, the property that its manifestations occur most notably when automation failures are extremely unexpected (e.g., truly surprising). These are what we describe as the "first failure effects" (Wickens, 2000, Yeh et al., 2003). Subsequent failures of automation now known by the supervisor to be imperfect appear to lead to less dramatic forms of attentional neglect of the automated process.

Statistical and Methodological Issues. The investigation of attentional tunneling is challenged by certain statistical issues. Most importantly, because it is an effect generally manifest with unexpected/ surprising events, it is a phenomenon that by definition can be effectively produced only one or two times per experiment (or per flight simulation). If the event used to document attentional tunneling occurs more frequently than this, it will by definition, no longer be surprising. One consequence of this fact is that pilot response to the event will be subject to high variability (since variability decreases with sample size, and the sample size will be small); as a consequence, the effects will be of relatively low statistical power, and researchers should be willing to accept a greater likelihood of committing a type 2 statistical error by raising their alpha level for significance above the 0.05 level (Wickens, 1996, 2001) when examining such responses to rare events.

(We note here the advantages of measuring visual scanning (Thomas & Wickens, 2004), a technique with relatively high statistical power, since it can be continuously measured, that can be a direct measure of attentional tunneling; thus a channel that is not looked at for a long period of time, can be inferred to produce neglect of important events that occur along that channel, should those events ever occur).

A methodological criticism that is sometimes directed toward the research typical of that above, which has documented attentional tunneling in flight simulation experiments, is that this is somewhat of an artificial phenomenon of the simulation laboratory, and that pilots flying in the "real world" would be more vigilant of such unexpected events, because of the higher stakes involved, and/or because of a greater expectancy that "anything can happen". On the one hand, there is some merit to this concern over generalizability. For example Fadden, Ververs and Wickens (2001) found that HUD-induced attentional tunneling was manifest for those pilots who had not participated in a flight simulation involving the off-normal runway incursion, but that the phenomenon was not shown by those who had previous experience. Thus it is possible that experience may mitigate the tunneling effect.

In response, however, two counterarguments can be given. First, the phenomenon has been demonstrated in very high fidelity simulations, by well qualified commercial pilots (Hofer et al., 2000). Second, higher levels of training may, ironically, make pilots less, rather than more likely to "expect the unexpected", if the unexpected event has never occurred within their many years of flight. A driving analogy is appropriate here. Most people drive on an expressway with a headway that is well less than the minimum to avoid a rear end collision should the leading driver suddenly come to a halt. This tendency is, in part, the result of never having experienced such an event.

Going beyond the issue of statistical and methodological issues, a strong case can be made that the safety implications of attentional tunneling may simply not be amenable to conventional statistical techniques that focus on "the statistics of the mean". This is because accidents, the target of generalization from our research, are not typical, and are probably not caused by human error of the "average" pilot flying in typical circumstances (Wickens, 2000, 2001). Rather, we might expect them to be caused by the poorly trained pilot, in high workload environments, perhaps, as noted above, dealing with a failure management scenario. Thus while only a small number of pilots may demonstrate the phenomenon of interest in the simulation laboratory, so also only a small number of pilots may demonstrate unsafe neglect and attentional tunneling in the sky in such a way as to lead to a mishap. Given that such accidents are well documented, any identification of factors that may invite greater tunneling, are worthy of empirical investigation. We hope that the factors discussed above contribute to that investigation.

Acknowledgments

This research was supported primarily by a grant from NASA Ames Research Center #NAG 2-308. Dr. David Foyle was the scientific/technical monitor.

References

Alexander, A. L., Wickens, C. D., & Hardy, T. J. (in press). Synthetic vision systems: The effects of guidance symbology, display size, and field of view. *Human Factors*.

Chou, C., Madhavan, D., & Funk, K. (1996). Studies of cockpit task management errors. *The International Journal of Aviation Psychology, 6*(4), 307-320.

Dismukes, K. (2001). The challenge of managing interruptions, distractions, and deferred tasks. *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus, OH: The Ohio State University.

Fadden, S., Ververs, P. M., & Wickens, C. D. (2001). Pathway HUDS: Are they viable? *Human Factors, 43*(2), 173-193.

Fischer, E., Haines, R. F., & Price, T. A. (1980). *Cognitive issues in head-up displays* (NASA Technical Paper 1711). Moffett Field, CA: NASA Ames Research Center.

Funk, K. H. (1991). Cockpit task management: Preliminary definitions, normative theory, error taxonomy, and design recommendations. *The International Journal of Aviation Psychology, 1*(4), 271-285.

Hockey, G. R. J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (vol. 2). New York: Wiley.

Hofer, E. F., Braune, R. J., Boucek, G. P., & Pfaff, T. A. (2000). *Attention switching between near and far domains: An exploratory study of pilots' attention switching with head-up and head-down tactical displays in simulated flight operations* (D6-36668). Seattle, WA: The Boeing Commercial Airplane Co.

Horrey, W.J. & Wickens, C.D. (in press). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors*.

Iani, C., & Wickens, C. D. (2004). Factors affecting task management in aviation. *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 213-217). Santa Monica, CA: HFES.

Moray, N. & Rotenberg, I. (1989). Fault management in process control: eye movements and action. *Ergonomics, 32*(11), 1319-1342.

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias: Decision making and performance in high-tech cockpits. *The International Journal of Aviation Psychology, 8*(1), 47-63.

Olmos, O., Wickens, C. D., & Chudy, A. (2000). Tactical displays for combat awareness: An examination of dimensionality and frame of reference concepts and the application of cognitive engineering. *International Journal of Aviation Psychology, 10*(3), 247-271.

Parasuraman, R. M., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation induced "complacency". *International Journal of Aviation Psychology, 3*, 1-23.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230-253.

Prinzel, L. J., III, Comstock, J. R., Jr., Glaab, L. J., Kramer, L. J., Arthur, J. J., & Barry, J. S. (2004). The efficacy of head-down and head-up synthetic vision display concepts for retro- and forward-fit of commercial aircraft. *The International Journal of Aviation Psychology*, *14*(1), 53-77.

Schnell, T., Kwon, Y., Merchant, S., & Etherington, T. (2004). Improved flight technical performance in flight decks equipped with synthetic vision information system displays. *The International Journal of Aviation Psychology*, *14*(1), 79-102.

Shappell, S. A., & Wiegmann, D. A. (2003). *A human error analysis of general aviation controlled flight into terrain accidents occurring between 1990-1998* (Final Rep. DOT/FAA/AM-03/4). Washington, DC: Office of Aerospace Medicine.

Strayer, D. L. & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on cellular telephone. *Psychological Science, 12*(6), 462-466.

Thomas, L. C., & Wickens, C. D. (2004). Eye-tracking and individual differences in off-normal event detection when flying with a synthetic vision system display. *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 223-227). Santa Monica, CA: HFES.

Wickens, C. D. (1996). Designing for stress. In J. Driskell & E. Salas (Eds.), *Stress and human performance* (pp. 279-295). Mahwah, NJ: Lawrence Erlbaum.

Wickens, C. D. (2000). *Imperfect and unreliable automation and its implications for attention allocation, information access and situation awareness* (Final Technical Report ARL-00-10/NASA-00-2). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Wickens, C. D. (2000). The tradeoff in the design for routine and unexpected performance. In M. Endsley & D. Garland (Eds.). *Situation awareness and measurement*. Mahwah, NJ: Lawrence Erlbaum.

Wickens, C. D. (2001). Keynote address: Attention to safety and the psychology of surprise. *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University.

Wickens, C. D., Alexander, A. L., Horrey, W. J., Nunes, A., & Hardy, T. J. (2004). Traffic and flight guidance depiction on a synthetic vision system display: The effects of clutter on performance and visual attention allocation. *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 218-222). Santa Monica, CA: HFES.

Wickens, C.D., Helleberg, J., & Xu, X. (2002). Pilot maneuver choice and workload in free flight. *Human Factors, 44*(2), 171-188.

Wickens, C. D., & Long, J. (1995). Object versus space-based models of visual attention: Implications for the design of head-up displays. *Journal of Experimental Psychology: Applied, 1*(3), 179-193.

Wickens, C. D., Ververs, P., & Fadden, S. (2004). Head-up display design. In D. Harris (Ed.), *Human factors for civil flight deck design* (pp. 103-140*)*. Ashgate.

Yeh, M., Merlo J. L., Wickens, C. D. & Brandenburg, D. L. (2003). Head up versus head down: The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling. *Human Factors, 45(3), 390-407.*

PILOT DEPENDENCE ON IMPERFECT DIAGNOSTIC AUTOMATION IN SIMULATED UAV FLIGHTS: AN ATTENTIONAL VISUAL SCANNING ANALYSIS.

Christopher Wickens, **Stephen Dixon**, **Juliana Goh** and **Ben Hammer**
University of Illinois, Aviation Human Factors Division
Savoy, Illinois

An unmanned air vehicle (UAV) simulation was designed to reveal the effects of imperfectly reliable diagnostic automation – a monitor of system health parameters – on pilot attention, as the latter was assessed via visual scanning. Four groups of participants flew a series of legs under different automation conditions: a baseline (no automation) control, and automation which was either 100% reliable, 60% reliable with a low-threshold bias to produce false alerts, and 60% reliable with a high threshold to produce misses. A high workload mission completion task and ground surveillance task were simultaneously imposed. Consistent with the reliance-compliance model of imperfect automation developed by Meyer (2001), miss-prone automation removed visual attention from the surveillance task, while FA-prone automation delayed the alert-driven attention shift to the system monitoring task.

Introduction

Unmanned air vehicles (UAV) have realized a recent successful history in military aviation, and presently are forecast to play an important role in civil aviation, either as military UAVs must transition through civilian airspace, or as UAVs are called upon to perform non-military functions such a border surveillance or cargo transport. UAVs, almost by definition, will require high levels of automation, and hence bring into play issues of pilot monitoring of that automation. Whether the pilot is called on to supervise a single UAV, as in most intended civilian applications, or two or more UAVs, as envisioned in many military applications, there are two major factors that mitigate the effectiveness of automation, in UAV control (as well as its effectiveness other aviation systems).

The first factor is the level of "**workload**" experienced by the human operator. Here we define workload, as the load imposed on the limited information processing resources of the unaided (without automation) human operator, in what we describe as the "baseline" or "manual" condition. This load can be imposed from two qualitatively distinct sources: the single task **difficulty** of the task that might otherwise be automated, and the **multi-task load** in which the baseline (vs. automated) task is performed. In these two cases, the automation benefits are likely to increase, to the extent that the single task to be automated is more difficult (Maltz & Shinar, 2003; Dixon & Wickens, 2004), or that concurrent or multi-task load is imposed (Parasuraman et al., 1993).

The second factor is automation **reliability**. There is little doubt that total human-system performance will be quite good if automation is perfect. Conversely, when performing a difficult task, performance will be poor when automation is so unreliable as to be useless. However in between these extremes, lies a range of reliability levels where the benefits of automation over the baseline may be uncertain.

Of course there are a wide array of types of automation that can be employed to assist the UAV pilot, as well as a wide variety of ways in which automation can fail. In the current research we focus on automated alerts, that are of particular value under high levels of pilot workload, because the attention-grabbing properties of such alerts typically relieve the pilot of continuous visual monitoring of the "raw data" in the "alerted domain". In our particular domain, the raw data represent indicators of the health of various systems on board the aircraft.

Three reasons lay behind our selection of this automated task for our research. First, because system monitoring is generally lower on the pilot's task Hierarchy (Schutte & Trujillo, 1996), it is logical to relegate this to an automated alert system. Secondly, interviews with subject matter experts of the Army's Hunter-Shadow UAV (Wickens & Dixon, 2002), revealed the plausibility of rendering such system failures as relatively frequent events, and therefore legitimate subjects of an experimental inquiry of imperfect automation. Finally, the nature of potential automated failures in monitoring system events generalizes to a much wider class of automated diagnostic systems in aviation, such as conflict and collision alerts (Bliss, 2003; Pritchett, 2001), so that lessons learned regarding the implications of this imperfect automation for pilot attention and decision, can be widely applied.

Underlying our current modeling approach is the fact that automated diagnostic systems must discriminate two kinds of events: a "failure" and a "normal operating condition". When asked to make such a discrimination in a probabilistic imperfect world, with potentially unreliable sensors, automation will make occasional errors. It is then the responsibility of the alert designer to "set the threshold" of the alerting system to achieve the appropriate balance of alert misses, and alert false alarms. Generally, designers have chosen to bias this setting in favor of a low threshold, which generates many more false alerts, than it does missed events (Pritchett, 2001); however, neither type of automation error is immune from human performance costs, imposed on the pilot who must (a) respond to the alert output (if it is true), (b) provide some attention to the "raw data" (to the extent that the alerting system may be miss-prone) and (c) perform a host of attention demanding concurrent tasks.

Some more specific description of what these costs are, emerges from a treatment of alert systems developed by Meyer (2001, 2004; Maltz & Shinar, 2003), who distinguishes between two cognitive states of human dependence on alerting automation: **Reliance**, characterizes human cognition when the alert is silent. A reliant operator will assume that the alert will unfailingly sound when the raw data go out of tolerance, and hence will have no need to examine those data while the alert is silent. Full residual attention will be available for concurrent tasks. However an imperfect alerting setting that generates automation misses will reduce reliance, at the expense of visual attention to concurrent tasks.

Compliance, in contrast, characterizes the operator response when the alert sounds. A highly compliant operator will rapidly abandon concurrent tasks and switch attention to the alerting domain once the alert sounds. However an imperfect alerting setting that generates many false alarms (the more frequent type of setting) will reduce compliance, even if this setting has minimal effect on reliance.

In a pair of UAV simulation experiments, Dixon and Wickens (2004; Dixon Wickens and Chang, 2005, in press) varied the auditory alerting threshold as well as the overall reliability of system monitor gauges in their simulated UAV. Examining performance on the system monitoring task itself, along with performance of a concurrent image surveillance task, and a primary mission task, they were able to demonstrate performance effects that appeared to mirror some of the expected changes in reliance and compliance: increasing automation miss rate reduced concurrent monitoring; increasing automation false alert rate reduced pilot response to system failures. Both of these effects reflect the inferred influence of automation reliability on **pilot attention**, either to monitor concurrent tasks, rather than the raw data (indexing high reliance), or to be immediately switched when an alert occurs (for a compliant pilot). however we had no direct measures of the allocation of visual attention, as revealed through visual scanning measures. Because of the critical role played by visual attention in aviation (Talleur & Wickens, 2003; Wickens, Goh, Helleberg, Horrey & Talleur, 2003), in the current study, we measured these scan patterns as four groups of pilots monitored simulations that varied in the reliability of the automated system status monitor: a 100% reliable system, an unreliable system ($r = 0.60$) with a bias to false alerts, an equally unreliable system ($r = 0.60$) with a bias to misses, and a baseline system with no auditory alerting whatsoever. In each system we measured performance, as well as the balance of visual attention between the system gauges and concurrent tasks (measuring miss-influenced reliance), and the visual attention switching time following an alert (measuring false-alert influenced compliance).

Methods

39 student pilots from the Institute of Aviation volunteered to participate in the experiment. They were paid $9.00/hour. Each pilot flew the UAV through ten different mission legs (one practice, 9 experimental), while completing three goal-oriented tasks commonly associated with UAV flight control: mission completion, target search, and systems monitoring. They used the interface shown in figure 1. At the beginning of each mission leg, pilots obtained flight instructions via the Message Box, including fly-to coordinates and a report question pertaining to the next command target (CT). These instructions were present for 15 seconds; in case the pilot forgot the instructions, pressing a repeat key refreshed the flight instructions for an additional 15 seconds.

Once pilots arrived at the CT location, they loitered around the target, manipulated a simulated camera for closer target inspection, and reported back relevant information to mission command (e.g., *What weapons are located on the south side of the building*?). This challenging CT report demanded motor, visual and cognitive resources (Gugerty & Brooks, 2001). Along each mission leg, pilots were also responsible for detecting and reporting low-salience targets of opportunity (TOO), a task similar to the CT report, except that the TOOs were much smaller (1-2 degrees of visual angle) and camouflaged. TOOs could occur during simple tracking (low workload) or during a pilot response to a system failure as described below (high workload).

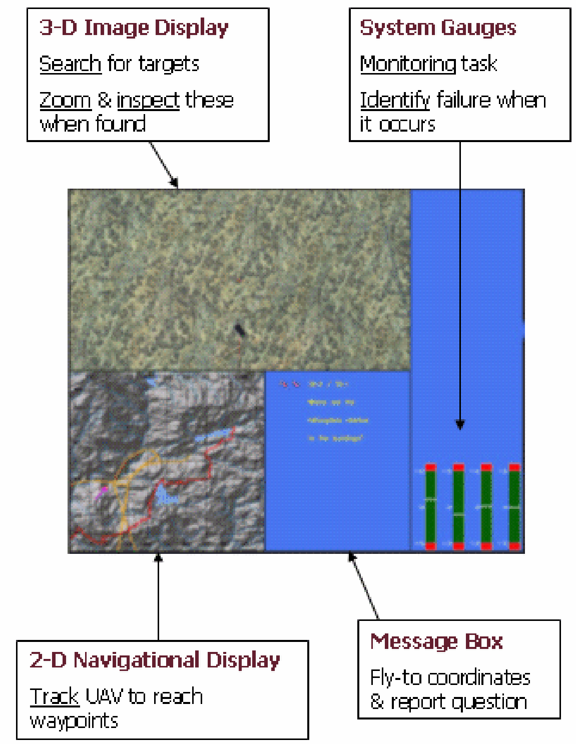


Figure 1: *The Experimental Display.*

Concurrently, pilots were also required to monitor the system gauges for possible system failures (SF). This was the "automated task". SFs were designed to fail either during simple tracking (low workload) or during TOO/CT inspection (high workload). The SFs lasted 30 seconds, after which the screen flashed bright red and a salient auditory alarm announced that the pilot had failed to detect the SF. There were a total of 10 SFs, with never more than two SFs occurring during any mission leg.

Automation aids, in the form of auditory auto-alerts during SFs, were provided for three of the four conditions. The **A100** condition (A = automation, 100% reliable) never failed to alert pilots of SFs. The **A60f** condition (f = false alarm, 60% reliable) failed by producing 3 false alarms and 1 miss out of the 10 SFs. The **A60m** condition (m = miss) failed by failing to notify pilots of a system failure on 3 of the 10 SFs, while generating a single false alarm. The final condition was a **baseline** manual condition , with no automation aid to assist pilot performance.

Pilots were not aware of the precise level of reliability of each automation aid; they were simply told that the automation was either "perfectly reliable" or "not perfectly reliable" and which way the threshold was set (i.e., whether the automation would produce false more false alarms or misses).

Results

Primary Task performance. The pilots' primary task was to fly the UAV to the command targets and make the report. Neither tracking accuracy nor CT report were much effected by automation reliability level, nor did this level effect pilot's memory for the CT information (as implicitly measured by the use of the "repeat" key). Hence pilots optimally protected this most important task from resource competition imposed by other tasks.

TOO monitoring. Prior studies had shown that this "secondary" task of monitoring the 3D image window was sensitive to the demands imposed by imperfection of the automation (Dixon & Wickens, 2004). Table 1a shows performance on the TOO task as a function of condition.

We focused our analysis on TOO responses that only occurred under low workload conditions, in which a system failure had not occurred (i.e., during the period of reliance) and observed the trend in both accuracy and speed to be degraded with less reliable automation, particularly in the miss-prone condition [although this trend was not significant for RT, and only marginally so for detection rate ($F3, 26 = 2.31$, $p=.10$)].

Table 1. *TOO and system failure monitoring/ detection performance.*

	Baseline	A100	A60F	A60M
(a) TOO				
(low workload)				
Acc (%)	89.00	82.00	75.00	61.00
RT (S)	6.05	6.50	7.64	10.10
(b) System Failure				
Low workload				
Acc	100.00	100.00	100.00	100.00
RT	7.19	2.18	3.02	3.36
High workload				
Acc	92.00	100.00	42.00	58.00
RT	11.46	4.82	23.28	14.77

System Failure Detection. Analysis of the system failure data revealed no effects of automation reliability at low workload, but that response times were faster when any type of automation was present, $F(3,26) = 5.40$, $p < .01$. Importantly, highly significant effects of reliability emerged at high workload, as revealed by the significant load X condition interaction in both Accuracy, $F(3,26)$ =

7.91, $p < .01$, and RT, $F(3,26) = 9.65$, $p < .01$. Our particular interest was in the differential cost between miss-prone and false alarm-prone automation, where, in the high workload condition, both accuracy (t=1.96, p=.04), and RT (t=3.53, p<.01) demonstrated a greater cost in the false-alarm prone condition than in the miss-prone condition.

Thus the emerging picture is one in which performance on both tasks suffers when automation reliability degrades, but SF performance degrades more severely, particularly in high workload, and with false-alarm prone automation, whereas monitoring of the 3D image window for TOOs degrades only slightly, and even then only in the miss-prone condition. Thus we now ask whether visual scanning behavior, a direct manifestation of attention allocation and switching, can provide any insight as to the role of reliance and compliance in mediating the above effects.

Visual Attention allocation. Table 2 provides a measure of the percent dwell time (PDT) that the eyes spent within each of the four areas of interest (AOI) on the workstation. The data are only reported during steady state (low workload) monitoring, not during the high workload segments involving zooming and panning of the 3D image window to identify detected targets. It is during this low workload period that pilots **rely** upon automation to alert them if such a system failure occurs.

Table 2. *Percentage Dwell Time that visual fixation is spent for the four experimental conditions within each area of interest (AOI): 3D image display where the TOOs were located, the 2D navigation display, the System failure monitoring gauges, and Message Box.*

AOI	Baseline	A100	A60F	A60M
3D (TOO)	50.0	58.7	56.4	45.5
2D	36.7	39.2	32.2	35.1
SF	13.0	5.7	11.3	18.6
MB	4.1	6.6	9.0	11.9

A 2 way (AOI X condition) ANOVA carried out on the PDT data revealed a significant effect of AOI, $F(3, 78) = 155.75$, $p < .001$. The 3D image window, hosting the most demanding surveillance and detection task demanded the most visual attention, the 2D nav display, hosting the most important task (command target location information) required around a third of the pilot's attention, and the two remaining AOIs demanded the least. Importantly, the significant AOI X condition interaction, $F(9,78) = 2.41$, $p = .05$, reflected automation reliance. Here we see that visual attention to the TOO window benefited (relative to baseline 50%) from having auditory alerts, whether these were fully reliable [100A, $t(14) = 2.05$, $p < .03$], or imperfect, but having few misses [60F, [$t(13) = 1.34$, $p = .10$]. However miss-prone automation drew as much, if not more visual attention away from the 3D window (45.5%) as this window received in the baseline condition (50%). While this decrease from the baseline was not significant, the difference between miss prone and false alarm prone automation was significant [t(13) = 1.7, p = .06], indicating the shift in attention to concurrent tasks, fostered by a designer's decision to change the alerting threshold.

Scanning to the 2D image display, hosting information for primary task navigation performance did not differ significantly between conditions, indicating how pilots treated this display which hosted primary task information, as of utmost priority. However scanning to the SF gauges themselves reflected an expected pattern, opposite to that of the 3D image window. While perfect automation (A100) greatly reduced the visual attention required, relative to baseline [$t(13) = 3.97$, $p<.01$], the miss-prone automation condition required far more visual attention to this display, as expected given that pilots are, presumably, paying more attention to the "raw data" compared to the false alarm prone condition [$t(13) = 2.05$, $p=.03$], which did not differ from baseline. An additional feature is that pilots paid even more attention (18%) in the miss-prone condition, than in the non-automated baseline (13%, $t = 1.71$, $p<.05$), a cost that, as we saw above, bought them nothing in terms of better SF detection performance. There was no difference in scanning to the message box across conditions.

One might not have expected the false alarm rate to influence reliance, and indeed it did not appear to influence the measures of the residual attention to the 3D image window where the TOOs appeared. However somewhat surprisingly, the higher FA rate did compel more attention to the SF display than the fully reliable automation condition, and induced no less attention there than the baseline condition. Thus no attention was "saved" by FA-prone automation relative to the baseline, in spite of the fact that nearly all failures were alerted. Thus, the general distrust induced by false alarms may have led to pilot suspicion that such a system requires further monitoring.

Visual Scan Response time. We inferred that compliance would be related to the speed with which visual attention moved to the SF gauges from wherever it was located at the time that the alert occurred. These measurements were computed by hand from a time-file of scanning across the 4 AOIs. The data for these "scan RT's" are shown in Table 3 when the alerts occurred during the high workload period while the pilot was engaged in image scanning:

Table 3. Scan RTs in seconds. (baseline scans represent the delay between the SF and the first look at the display. All others represent the delay between the auditory alert and the first look).

Baseline	A100	A60F	A60M
19s	4.5s	16 s	4.0s

A one way ANOVA on these data revealed a highly significant effect of condition, $F(3,29) = 5.806$, $p = .004$, revealing that looks were as rapid in the miss-prone condition, as in the perfect automation condition (pilots' perfectly complying with the alerts), but were as slow in the false-alarm condition as were the unaided glance times.

Discussion

The current results extended the previous findings of imperfect diagnostic automation in UAVs (Dixon & Wickens, in press) to consider the explicit response of pilot attention, underlying the two inferred constructs of reliance and compliance. These two constructs characterize a pilot's response to automation that has a low miss rate and a low false alarm rate respectively.

As in the previous study, we found that an increasing miss rate produced a marginal loss in concurrent task performance. In the current data we noted that this was paralleled (and presumably caused) by a re-allocation of visual attention away from the 3D image window, toward the raw data hosted within the SF display (i.e., toward the oscillating bars representing system parameter health).

Also as in the previous study, we found that an increasing automation false alert rate, while having little effect on concurrent task performance (or attention allocation to the concurrent task), yielded a pronounced loss in SF detection performance in high workload, causing misses of some true alerts, and substantial delays in responding to all alerts. Interestingly, the increase in mean response time from the perfect automation condition to the A60F condition was 19 sec (Table 1b), whereas the increase in mean scan RT was only 11.5 sec (Table 3). Such a difference indicates that, when false alarm rate was high, alert-driven looks to the display were followed by an additional 7.5 seconds of examining the raw data to assure that the alert was a true one, before an overt response was given. Overall, this delay, reflecting the cost of false-alarm prone automation, is of significant duration to be of significant operational importance.

The current data reinforces the notion that imperfect automation effects can be well modeled by their influence on pilot attention, and that such effects can be profound if automation reliability is allowed to drop to levels of around 60%, well below the threshold of approximately 70% reliability revealed to determine when automation is no longer useful (Wickens & Dixon, 2005). While such rates may seem, at first glance, to be unrealistically low, it should be noted that in many aviation circumstances diagnostic automation is asked to **predict** events in a probabilistic world, plagued by future uncertainties in such variables as human response, or turbulence (Xu, Rantanen & Wickens, 2005; Thomas, Wickens & Rantanen, 2003; Krois, 1999). Under such circumstances, reliability rates not unlike those examined here, may be expected. It is therefore important that the consequences of these rates to pilot/supervisor performance are well understood.

Acknowledgments

This research was sponsored by subcontract #ARMY MAD 6021.000-01 from Microanalysis and Design, as part of the Army Human Engineering Laboratory Robotics CTA, contracted to General Dynamics. David Dahn was the scientific/technical monitor. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the Army. The authors also wish to acknowledge the support of Ron Carbonari and Jonathan Sivier (in developing the UAV simulation).

References

Bliss, J. (2003). An investigation of alarm related accidents and incidents in aviation. *International Journal of Aviation Psychology, 13*(3), 249-268.

Dixon, S. R., & Wickens, C. D. (in press). Automation reliability in unmanned aerial vehicle flight control: Evaluating a model of automation dependence in high workload. *Human Factors.*

Dixon, S. R., Wickens, C. D., & Chang, D. (2005, in press). Mission control of unmanned air vehicles: A workload analysis. *Human Factors, 47.*

Gugerty, L., & Brooks, J. (2001). Seeing where you are heading: Integrating environmental and egocentric reference frames in cardinal direction judgments. *Journal of Experimental Psychology: Applied, 7*(3), 251-266.

Krois, P. (1999, July 25). *White Paper: Human factors assessment of the URET conflict probe false alert rate*. Washington, DC: Federal Aviation Administration.

Maltz, M., & Shinar, D. (2003). New alternative methods of analyzing human behavior in cued target acquisition. *Human Factors, 45*(2), 281-295.

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors, 43*, 563-572.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors.*

Parasuraman, R. M., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation induced "complacency". *International Journal of Aviation Psychology, 3*, 1-23.

Pritchett, A. (2001). Reviewing the role of cockpit alerting systems. *Human Factors & Aerospace Safety, 1*, 5-38.

Schutte, P. C., & Trujillo, A. C. (1996). Flight crew task management in non-normal situations. *Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 244-248). Santa Monica, CA: HFES.

Talleur, D.A., & Wickens, C.D. (2003). The effect of pilot visual scanning strategies on traffic detection accuracy and aircraft control. *Proceedings of the 12th International Symposium on Aviation Psychology*. Dayton, OH: Wright State University.

Thomas, L.C. Wickens, C.D., & Rantanen, E.M. (2003). Imperfect automation in aviation traffic alerts: A review of conflict detection algorithms and their implications for human factors research. *Proceedings of the 47th Annual Meeting of the Human Factors & Ergonomics Society*. Santa Monica, CA: HFES.

Wickens, C. D., & Dixon, S. (2002). *Workload demands of remotely piloted vehicle supervision and control: (I) Single vehicle performance* (ARL-02-10/MAD-02-1). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Wickens, C. D., & Dixon, S. (2005). Is there a magic number 7 (to the minus 1)? The benefits of imperfect diagnostic automation: A synthesis of the literature (AHFD-05-1/MAAD-05-1). Savoy, IL: University of Illinois, Aviation Human Factors Division.

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W., & Talleur, D. A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors, 45*(3), 360-380.

Xu, X., Rantanen, E. & Wickens, C. D. (2005). Effects of conflict warning system reliability and task difficulty on pilots' conflict detection with cockpit display of traffic information. *Proceedings of the International Symposium on Aviation Psychology*. Dayton, OH: Wright State University.

A SUMMARY OF UNMANNED AIRCRAFT ACCIDENT/INCIDENT DATA: HUMAN FACTORS IMPLICATIONS

Kevin W. Williams, Ph.D.
FAA Civil Aerospace Medical Institute
Oklahoma City, OK

A review and analysis of unmanned aircraft (UA) accident data was conducted to identify important human factors issues related to their use. UA accident data were collected from the U.S. Army, Navy, and Air Force. The percentage of involvement of human factors issues varied across aircraft from 21% to 68%. For most of the aircraft systems, electromechanical failure was more of a causal factor than human error. One critical finding from an analysis of the data is that each of the fielded systems is very different, leading to different kinds of accidents and different human factors issues. A second finding is that many of the accidents that have occurred could have been anticipated through an analysis of the user interfaces employed and procedures implemented for their use. The current paper summarizes the various human factors issues related to the accidents

Introduction

The review and analysis of unmanned aircraft (UA) accident data can assist researchers in identifying important human factors issues related to their use. The most reliable source for UA accident data currently is the military. The military has a relatively long history of UA use and has always been diligent in accurately recording information pertaining to accidents/incidents. The purpose of this research was to review all currently available information on UA accidents and identify human error aspects in those accidents and what human factors issues are involved.

Two primary sources of accident information were collected from the U.S. Army. The first was a summary of 56 UA accidents produced by the U.S. Army Aeromedical Research Laboratory (Manning, Rash, LeDuc, Noback, & McKeon, 2004) and obtained from the U.S. Army Risk Management Information System (RMIS). The second was a direct query of the RMIS system of all UA accidents that occurred between January 1986 and June 2004. A total of 74 accidents were identified, the earliest of which occurred on March 2, 1989, and the latest on April 30, 2004.

Information regarding UA accidents for the U.S. Navy was collected from the Naval Safety Center. A summary of 239 UA mishaps occurring between 1986 and 2002 was received from the Naval Safety Center in Pensacola, FL (Kordeen Kor, personal communication).

Air Force accident/mishap information was collected from the Air Force Judge Advocate General's Corps Web site, http://usaf.aib.law.af.mil/. A total of 15 Class-A UA mishaps were retrieved from the Web site, covering the dates from December 6, 1999, to December 11, 2003. In addition, a complete accident investigation board report was received.

Classification of the accident data was a two-step process. In the first step, accidents were classified into the categories of human factors, maintenance, aircraft, and unknown. Accidents could be classified into more than one category. In the second step, those accidents classified as human-factors-related were classified according to specific human factors issues of alerts/alarms, display design, procedural error, skill-based error, or other. Classification was based on the stated causal factors in the reports, the opinion of safety center personnel, and personal judgment of the author.

Results

There are 5 primary military UA in service currently. The U.S. Army's Hunter and Shadow, the U.S. Navy's Pioneer, and the U. S. Air Force's Predator and Global Hawk. Other systems are being developed and have undergone testing, such as the Mariner system for the U.S. Coast Guard and U.S. Navy but sufficient accident data do not exist to warrant separate analyses of these airframes.

Hunter

The Hunter takes off and lands using an external pilot (EP), standing next to the runway in visual contact with the aircraft, and operating a controller that is very similar to ones used by radio-controlled aircraft hobbyists. After takeoff and climb out, control of the aircraft is transferred to an internal pilot (IP), operating from a ground control station (GCS). The IP controls the Hunter in a more automated fashion, by selecting an altitude, heading, and airspeed for the aircraft using a set of knobs located within the GCS. For landing, control of the aircraft is transferred from the GCS back to an EP. A hook located below the aircraft is used to snag the aircraft on a set of arresting cables positioned across the runway.

Data from the Hunter program indicated that 15 of the 32 accidents (47%) had one or more human factors issues associated with them. Figure 1 shows the major causal categories for Hunter accidents. Note that the percentages add to more than 100% because some of the accidents were classified into more than one category.

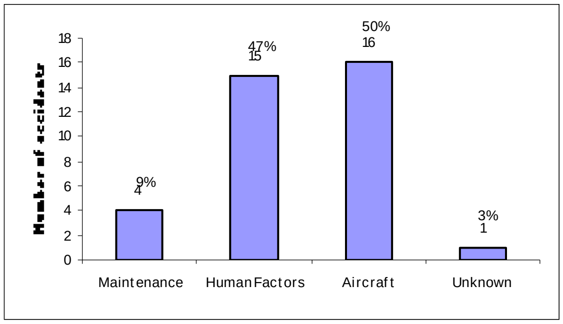


Figure 1. *U.S. Army Hunter accident causal factors.*

Breaking down the human factors issues further, Table 1 shows how the number and percentage of the 15 human-factors-related accidents are associated with specific human factors issues. Again, percentages exceed 100% because of some accidents being classified under more than one issue.

Table 1. *Breakdown of human factors issues for Hunter accidents.*

Issue	Number	Percent
Pilot-in-command	1	7%
Alerts and Alarms	2	13%
Display Design	1	7%
External Pilot Landing Error	7	47%
External Pilot Takeoff Error	3	20%
Procedural Error	3	20%

By far the largest human factors issue is the difficulty experienced by EPs during landings. Forty-seven percent of the human factors-related Hunter accidents involved an error by the EP during landing. An additional 20% of the accidents involved an error by the EP during takeoff. Control difficulties are at least partially explainable by the fact that when the aircraft is approaching the EP the control inputs to maneuver the aircraft left and right are opposite what they would be when the aircraft is moving away from the EP. This cross-control problem is present for any UA operated by an external pilot via visual contact.

Besides EP control problems, other issues represented in the table include pilot-in-command issues, alerts and alarms, display design, and crew procedural error. A pilot-in-command issue is a situation where the authority of the controlling pilot is superceded by other personnel in the area, violating the principle that the pilot of the aircraft has the final decision-making authority during a flight. In contrast, alerts and alarms deal with situations where a non-normal flight condition (e.g., high engine temperature) is not conveyed effectively to the crew. Display design issues typically manifest when not all of the information required for safe flight is conveyed effectively to the crew.

Finally, the crew procedural errors referred to here involved three occasions where the crew failed to properly follow established procedures. On one occasion an improper start-up sequence led to data link interference from the backup GCS. On another occasion the crew failed to follow standard departure procedures and the UA impacted a mountain. On a third occasion an EP failed to complete control box checks prior to taking control of the UA and did not verify a box switch that was in the wrong position.

Shadow

Unlike the Hunter, the Shadow (see Figure 2) does not use an external pilot, depending instead on a launcher for takeoffs, and an automated landing system for recovery. The landing system, called the tactical automated landing system (TALS) controls the aircraft during approach and landing, usually without intervention from the GCS pilot. A cable system, similar to the one used for the Hunter, is used to stop the aircraft after landing. Aircraft control during flight is accomplished by the GCS pilot through a computer menu interface that allows

Figure 2. *U.S. Army Shadow*

selection of altitude, heading, and airspeed. During landing, GCS personnel have no visual contact with the aircraft, nor do they have any sensor input from onboard sensors. A command to stop the aircraft engine is given by the GCS pilot, who must rely on an external observer to communicate that the plane has touched down.

The analysis of Shadow accidents shows a different pattern from that seen with the Hunter. In contrast to the Hunter, only 5 of the 24 Shadow accidents (21%) were attributed to human factors issues. Figure 3 shows the major causal factors for the Shadow accidents.

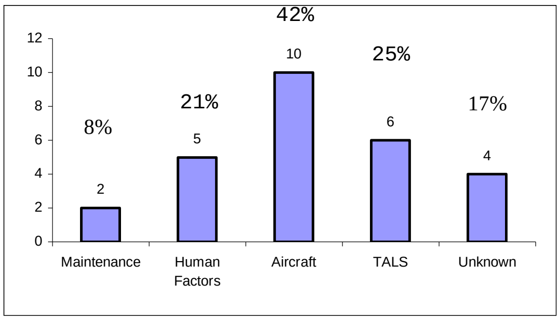


Figure 3. *U.S. Army Shadow accident causal factors.*

In addition to the four categories used for the Hunter accidents, an additional category was added for Shadow to include failures of the tactical automated landing system (TALS). While eliminating landing accidents potentially attributable to an EP, the use of TALS is not perfect, as shown from the data. Use of the launcher eliminated any EP takeoff errors for these aircraft.

Breaking down the human-factors-related accidents, Table 2 shows the number and percentage of the 5 accidents related to specific human factors issues. As can be seen from the table, the distribution of issues is evenly divided across pilot-in-command, alerts and alarms, display design, and procedural errors. The percentages sum to greater than 100% because of multiple attributions for some accidents.

Table 2. *Breakdown of human factors issues for Shadow accidents.*

Issue	Number	Percent
Pilot-in-command	2	40%
Alerts & Alarms	2	40%
Display Design	2	40%
Procedural error	2	40%

For both the Hunter and Shadow, at least one accident involved the transfer of control of the aircraft from one GCS to another during flight, an activity unique to UA. In the case of the Shadow, two aircraft were damaged during a single mission. The first was damaged due to a TALS failure. After the accident, the GCS crew issued a command to the damaged aircraft to kill its engine, but because of damage to the antenna the command was not received. That same GCS was then tasked with controlling a second Shadow that was on an approach. Unfortunately, after taking control of the second Shadow, the aircraft received the "engine kill" command that was still waiting for an acknowledgment from the GCS software, causing the second Shadow to crash also. This accident was classified as both a procedural error, because the crew failed to follow all checklist items prior to the transfer of control of the second aircraft, and a display design problem, because there was not a clear indication to the crew of the status of the "engine kill" command that had been issued.

Pioneer

Like the U.S. Army's Hunter UA, the Pioneer requires an EP for takeoff and landing. After takeoff, the aircraft can be controlled from a GCS in one of three modes. In the first mode the air vehicle is operated autonomously and the autopilot uses global positioning system (GPS) preprogrammed coordinates to fly the air vehicle to each waypoint. In the second mode, the IP commands the autopilot by setting knobs (rotary position switches) to command airspeed, altitude, compass heading or roll angle, and the autopilot flies the UA. In the third mode, the IP flies the aircraft using a joystick. The Pioneer can be landed at a runway using arresting cables, but because it is a U.S. Navy/Marine operated aircraft, it is also landed on board a ship by flying into a net. There are plans for implementing an automated landing system for the Pioneer for ship-based landings.

A list of 239 Pioneer accidents was received from the Navy Safety Center. Although not providing much detail, the data did allow a general categorization of accidents into principle causal categories. Figure 4 shows the major causal factors for Pioneer accidents.

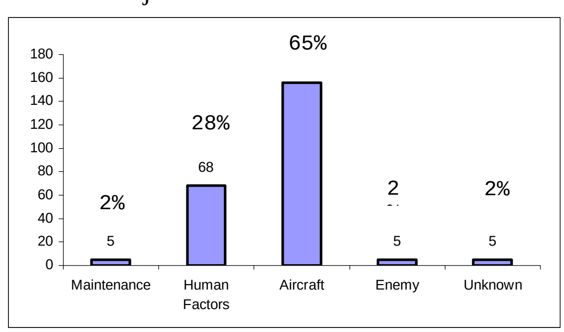


Figure 4. *U.S. Navy Pioneer UA accident causal factors.*

As can be seen from the figure, human factors-related issues were present in approximately 28% of the accidents. Breaking down the human factors-related accidents further, Table 3 lists the number and percentage of the 68 accidents related to specific human factors issues.

Table 3. *Breakdown of human factors issues for Pioneer accidents.*

Issue	Number	Percentage
Aircrew Coordination	9	13%
Landing Error	46	68%
Take-off Error	7	10%
Weather	6	9%

As with the U.S. Army Hunter accidents, the largest percentage of human factors accidents (68%) was associated with the difficulty experienced by the EP while landing the aircraft. An additional 10% of the accidents were associated with takeoffs, although the primary means of taking off is through the use of a launcher (from ship-based aircraft). In addition to landing and takeoff errors, two other issues seen with the Pioneer were aircrew coordination, which includes procedural and communication type errors, and weather-related accidents, which deal with pilot decision-making. Unfortunately, details regarding these accidents were not sufficient to identify issues beyond this level.

Predator

The Predator made its first flight in June 1994. There are two Predator types, currently designated as MQ-1 and MQ-9, also called Predator and Predator B. The Predator aircraft is flown from within the GCS, similarly to a manned aircraft, using a joystick and rudder pedals and a forward-looking camera that provides the pilot with a 30-degree field of view. The camera is used for both takeoffs and landings.

The Predator accident causal factors are shown in Figure 5. As can be seen from the figure, human factors encompass a higher percentage (67%) than aircraft-related causes, unlike the other aircraft examined thus far.

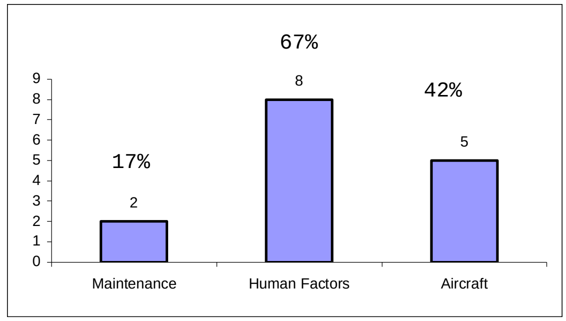


Figure 5. *Air Force Predator accident causal factors.*

Table 4 shows a breakdown of the human factors issues associated with Predator accidents. The majority of human-factors-related problems were concerned with procedural errors on the part of the flight crew. One of these accidents involved yet another problem with a handoff of the aircraft from one GCS to another. During the handoff, the mishap crew did not accomplish all of the checklist steps in the proper order, resulting in turning off both the engine and the stability augmentation system of the aircraft. The aircraft immediately entered an uncommanded dive and crashed.

Table 4. *Breakdown of human factors issues for Predator accidents.*

Issue	Number	Percentage
Alerts & Alarms	1	13%
Display Design	2	25%
Landing Error	1	13%
Procedural Error	6	75%

A second procedural error of note occurred when the pilot accidentally activated a program that erased the internal random access memory on board the aircraft during a flight. That this was even possible to do during a flight is notable in itself and suggests the relatively ad hoc software development process occurring for these systems (Tvaryanas, 2004).

Global Hawk

The Global Hawk is the largest and newest of the 5 military systems discussed. The first flight of the Global Hawk occurred in February 1998, and it became the first UA to cross the Pacific Ocean in April 2001 when it flew from the United States to Australia (Schaefer, 2003).

The Global Hawk is the most automated of all the systems discussed. All portions of the flight, including landing and takeoff are pre-programmed before the flight and the basic task of the crew during the flight is simply to monitor the status of the aircraft and control the payload. While this makes flying the Global Hawk very simple, the mission planning process is unwieldy and requires a great deal of time to accomplish.

Only three accident reports were available for the Global Hawk. Of these three reports, one did not provide sufficient information for classification, a second faulted a failure in a fuel nozzle, which led to an engine failure, and the third was a human factors issue centering on the complicated mission planning process. In that accident, the mishap aircraft suffered an inflight problem with temperature regulation of the avionics compartment and landed at a preprogrammed alternate airport for servicing. After landing, the aircraft was commanded to begin taxiing. Unknown to the crew, a taxi speed of 155 knots had been input into the mission plan at that particular waypoint as a result of a software bug in the automated mission planning software in use at the time. The aircraft accelerated to the point where it was unable to negotiate a turn and ran off of the runway, collapsing the nose gear and causing extensive damage to the aircraft.

Conclusions

One conclusion apparent from the data reported here is that, for most of the systems examined, electrical and mechanical reliability play as much or more of a role in the accidents as human error. Mishaps attributed at least partially to aircraft failures range from 33% (Global Hawk) to 67% (Shadow) in the data reported here.

An improvement in electromechanical reliability will probably come only through an increase in the cost of the aircraft. However, a reduction of human errors leading to accidents might not necessarily entail increased costs if suggested changes can be incorporated early in the design process. In the systems analyzed, human factors issues were present in 21% (Shadow) to 67% (Predator) of the accidents. These numbers suggest there is room for improvement if specific human factors issues can be identified and addressed.

In that regard, it is important to note that many of the human factors issues identified are very much dependent on the particular systems being flown. For example, both the Pioneer and Hunter systems have problems associated with the difficulty external pilots have in controlling the aircraft. For both of these systems, the majority of accidents due to human error can be attributed to this problem. However, the other three systems discussed do not use an EP and either use an IP (Predator) or perform landings using an automated system (Shadow and Global Hawk).

The designs of the user interfaces of these systems are, for the most part, not based on previously established aviation display concepts. Part of the cause for this is that the developers of these system interfaces are not primarily aircraft manufacturers. Another reason is that these aircraft are not "flown" in the traditional sense of the word. Only one of the aircraft reviewed (Predator) has a pilot/operator interface that could be considered similar to a manned aircraft. For the other UA, control of the aircraft by the GCS pilot/operator is accomplished indirectly through the use of menu selections, dedicated knobs, or preprogrammed routes. These aircraft are not flown but "commanded." This is a paradigm shift that must be understood if appropriate decisions are to be made regarding pilot/operator qualifications, display requirements, and critical human factors issues to be addressed.

If the aircraft is commanded to begin taxiing, there should be information available regarding the intended taxi speed. If the aircraft is being handed off from one station to another, the receiving station personnel should be aware of what commands will be transmitted to the aircraft after control is established. Interface development needs to be focused around the task of the pilot/operator. For most of these aircraft, that task is one of issuing commands and verifying that those commands are accepted and followed. Understanding this task and creating the interface to support it should help to improve the usability of the interface and reduce the number of accidents for these aircraft. This is especially important as these aircraft begin to transition to the National Airspace System (NAS), conducting civilian operations among civilian manned aircraft.

References

Manning, S.D., Rash, C.E., LeDuc, P.A, Noback, R.K., & McKeon, J. (2004). The role of human causal factors in U.S. Army unmanned aerial vehicle accidents. U.S. Army Aeromedical Research Laboratory Report # 2004-11.

Schaefer, R. (2003). Unmanned aerial vehicle reliability study. Office of the Secretary of Defense, Washington, DC, February 2003.

Tvaryanas, A.P. (2004). USAF UAV mishap epidemiology, 1997-2003. Presented at the *Human Factors of Uninhabited Aerial Vehicles First Annual Workshop*, Phoenix, AZ, May 24-25, 2004.

A REAL-TIME AUTOMATED TRANSCRIPTION TOOL FOR TACTICAL COMMUNICATIONS ASSESSMENT

David T. Williamson
Robin A. Snyder Jr.
Human Effectiveness Directorate
Wright-Patterson AFB, Ohio

1Lt Mathew D. Purtee
Human Effectiveness Directorate
Mesa, Arizona

This paper describes the development of a real-time automated transcription tool for assessing tactical communications in a DIS environment. Java-based tools were developed to capture simulated radio communications data from tactical training exercises conducted at the Warfighter Training Research Division of the Air Force Research Laboratory's Human Effectiveness Directorate. A representative set of audio data was hand-transcribed and used as training material for a class-based statistical language model using a commercially available speech recognition system. The language model was designed to allow run-time input of callsign data to provide increased flexibility. The resulting system is a real-time automated speech-to-text transcription tool that logs the audio data obtained from signal PDUs as a standard wave file and produces a text transcription, aiding in assessing tactical communications effectiveness. To test the capabilities of this system, an evaluation was performed using DIS log data from similar training exercises. Preliminary results indicate that overall word error rates across all participants were around 18%. For individual stations, however, word error rates less than 4% were obtained. Additional efforts are underway to refine the language model to achieve further reductions in error rate. Also discussed are efforts to further develop the Java DIS tools to provide scanner, logger, and basic radio functionality.

Background

Scientists at the Air Force Research Laboratory (AFRL) are involved in an effort to provide technological advances that will pay off for information management and training development for years to come. According to Version 6 of the Department of Defense Joint Technical Architecture (2003),

"For US forces to counter current and future threats successfully, they must operate worldwide with speed, agility, and flexibility. Key to achieving this required level of responsiveness is providing the quality, share situational awareness, and understanding necessary to make sound individual and collective judgments."

The emphasis of this doctrine is for continual development and maximum training of situational awareness and communications with team members from different backgrounds. Currently, there is a need for automatic scoring of radio communication during Distributed Mission Operation (DMO) events and training system applications for Air Force linguists (Air Force Policy Directive 33-1, 2001). In order to accomplish this, AFRL is evaluating available Speech-To-Text (STT) applications and incorporating Latent Semantic Analysis (LSA) in order to develop, demonstrate, and integrate in near real time automatic speech evaluation capabilities in training disciplines. Another goal is the automatic ability to transfer languages other than English in communication amongst team members.

Known as the Warfighter Communications Assessment System (WCAS) (Figure 1), this effort will provide Air Combat Command (ACC) with a comprehensive speech recognition, database, and analysis capability that will be instrumental to future readiness assessments and training delivery. This will be accomplished by assessing the feasibility of automated communications evaluation for training and demonstrating the capability of using standard scores as a criterion for training and rehearsal. Given success of this developing technology, automatic scoring will provide objective training effectiveness and proof of concept retention on paper as well as provide capabilities to communicate in languages other than English.

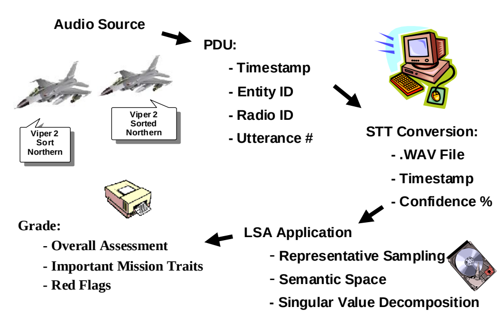


Figure 1. *Warfighter Communications Assessment System (WCAS)*

The first applicable domain of WCAS will be the F-16 four ship scenario. This need is outlined in Air Force Instruction (AFI) 11-2F-16, Flying Operations (2002):

"Units will design training programs to achieve the highest degree of combat readiness consistent with flight safety and resources availability. Training must balance the need for realism against the expected threat, pilot capabilities, and safety."

Benefits from this advanced training tool will most likely include enhanced mission effectiveness, reduced fratricide, enhanced training needs, and improved training. A highly proficient cadre of operators for these domains will be more readily available for current and future mission needs.

Directed by AFRL/HEAS (Warfighter Skill Division and Training Branch) and supported by efforts from HEC (Warfighter Interface Division) and KAT (Knowledge Analysis Technologies), this concept will be achieved through a number of significant milestones. The first is to develop an intelligent information retrieval workstation with a speech recognition system integrated and then fed into a LSA tool – an advanced statistical algorithm methodology - used to score aircraft communications. The second milestone is to assess the impact of less-than-ideal STT data basing on the accuracy of embedded content assessment, data flagging, and monitoring. These findings will then be developed for demonstration and integration of the WCAS capabilities into other key operational settings such as tactical communications, Battlestaff commander action planning and decision making, and information warfare. The final effort of the program is to demonstrate the transferability of the tools, methods, and data to languages other than English.

Latent Semantic Analysis is a mathematical tool for evaluating the contextual-usage meaning of words by means of statistical computations (Landauer, Foltz, & Laham, 1998). The WCAS program was first based upon an LSA-based tool for tracking and scoring text through research AFRL sponsored as part of a Small Business Innovative Research effort (Laham, Bennett, & Derr, 2001). This work demonstrated very basic speech recognition and LSA analysis capabilities which validated basic concepts and highlighted applications, such as career field and information operations applications, and discussed challenges with real time processing and further research this current effort will address.

This paper describes the first two steps in the overall WCAS implementation: the audio extraction and logging of DIS signal PDUs and the subsequent recognition and transcription of these communication events. The design and implementation of the extraction and transcription processes will be described along with the results of evaluations on sample scenario data. Also discussed are efforts to further develop the Java-based DIS tools to provide scanner and basic radio functionality.

Language Model Development

One of the features of the speech recognition system is the ability to use statistical language models (SLMs) to represent the target domain. This SLM technique was used in a recent study to assess performance on a NATO Native and Non-Native (N4) speech database (Williamson & Snyder, 2002). This approach, combined with a feature called robust natural language interpretation, provides a powerful capability to recognize a wide variety of commands. The steps required in creating an SLM grammar are 1) create the training set; 2) optionally create a vocabulary file, 3) determine the order of the SLM, 4) train the SLM, and 5) incorporate the SLM in the application. These steps applied to the WCAS domain are described below.

Training Set

The first step in generating an SLM was the creation of the training set. This involved the manual transcription of a set of DMO training sessions that represented the potential range of communications events that were likely to occur in this domain. Seven DMO sessions were used for the training set development. To maximize the effectiveness of the training set, a number of grammar rules or classes were created for those vocabulary items that were likely to change from one session to the next. Two of these classes were Fighter and AWACS callsigns. By substituting specific callsigns in the training data with generic placeholders for Fighter and AWACS, the resulting SLM is able to recognize any callsigns provided at runtime. Other classes included items such as Heading, Altitude and Range values.

Vocabulary File

The next step was the generation of the vocabulary list. While this step was optional, it provided the ability to constrain the model to only those vocabulary items relevant to the domain, excluding items such as word fragments or other disfluencies.

SLM Order

Next, the order of the SLM was determined. The order of an SLM refers to the probability assigned to words that occur in groups of N. In a bigram model, this probability is calculated for groups of two words. To find the optimum value of N for the n-gram language model, several recognition experiments were conducted varying the order from a bigram to a trigram model. In the end, the trigram model was chosen as providing the best balance of performance with overall model complexity.

SLM Training

The next step was training the SLM. This was done using a utility that takes the training set, vocabulary file and model order as inputs and provides a resulting file that can be incorporated into the application grammar.

Application Grammar

The final step in the overall language modeling process was the creation of the application grammar. Here, specific callsigns are inserted dynamically prior to a given data collection session. The application grammar also provided for semantic tagging of specific data items, such as AWACS and fighter callsigns, which were returned by the robust natural language interpretation engine upon recognition of a given utterance.

Language Model Evaluation

To test the performance of the language model, an evaluation was performed on the six remaining session logs that were not used in training the model. Figure 2 shows the average word error rates across the five players, AWACS and Fighters 1-4. This overall error rate combines substitution errors, where an incorrect word is substituted for the correct word, deletions, where a word is spoken but not returned in the recognition result, and insertions, where additional words are incorrectly inserted in the result when the speaker did not speak them.

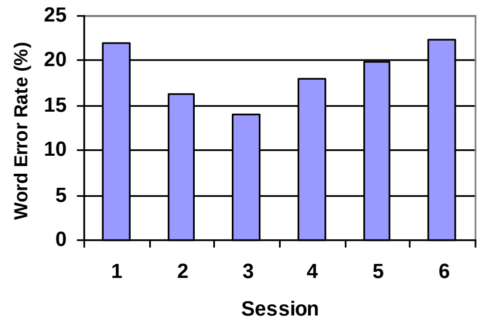


Figure 2. *Overall Word Error Rates by Session*

The performance of the language model on the individual station positions for each of the six test sessions is shown in Table 1.

Table 1. *Individual Word Error Rates by Session by Station*

Session	AWACS	F1	F2	F3	F4
1	11.1	21.4	22.9	25.2	29.3
2	3.6	30.7	32.9	32.0	24.1
3	5.9	26.0	22.0	13.4	24.1
4	9.3	24.9	23.8	23.8	26.9
5	10.3	37.3	19.7	21.7	27.5
6	11.0	34.1	26.3	28.3	27.7

Figure 3 shows the average word error rate across the six test sessions for each of the five stations. Clearly, the AWACS station achieved the best overall performance. AWACS also represented the greatest percentage of total words spoken with 44.6%, followed by Fighter 1 (20.8%), Fighter 3 (17.1%), Fighter 2 (20.8%), and Fighter 4 (7.9%).

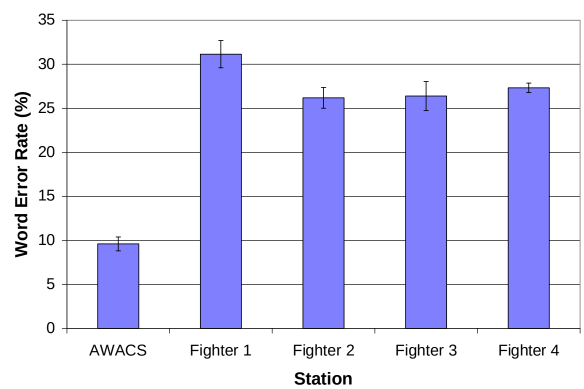


Figure 3. *Average Word Error Rates by Station*

Transcription Tool Development

The transcription tool is broken into three parts, the Courier, Transcriber, and Recognizer and is depicted in Figure 4. Each part has its own unique responsibilities and is described below.

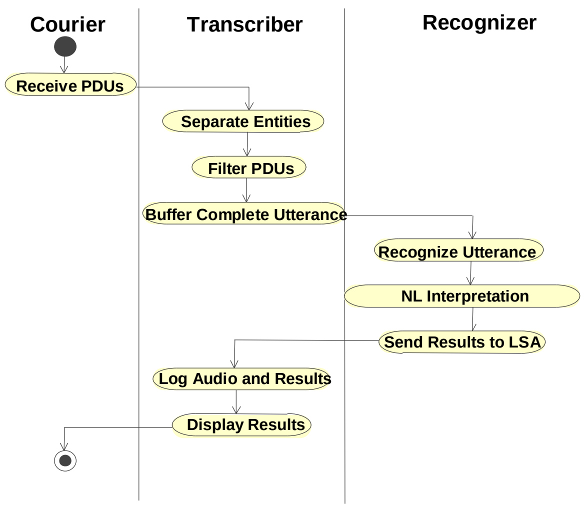


Figure 4. *Activity Diagram*

Courier

The Courier is the interface to the IP network. Its sole responsibility is to send and receive Protocol Data Units (PDUs). The Courier listens on port 3000, as specified in the IEEE standard, for UDP network traffic. All UDP packets are received and analyzed to determine if they are PDUs. If the packet is found to be a PDU, it is then sent to the Transcriber.

Transcriber

The Transcriber processes the PDUs, preparing them for recognition, and then logs and displays the results. The first step is to separate the PDUs based on their Entity ID and Radio ID placing the PDU into its entities sorted list. When the PDUs are added to the list, they are filtered and sorted. The PDUs are filtered according to type and timestamp. The filter first checks to see if the PDU is either a signal or transmitter, as these are the only PDUs necessary for audio extraction. If the PDU is of these types, the timestamp is then checked to be within the time window of the PDUs contained in the sorted list. The list is sorted by timestamp and then by type. A transmitter that specifies the radio is on and transmitting has higher priority than a signal. Transmitters with the other transmit states, off and on and not transmitting, have lower priority than signals. A complete utterance is made up of the following PDU sequence: a transmitter PDU indicating on and transmitting, a series of signal PDUs containing audio data and a transmitter PDU indicating the radio is not transmitting. This utterance is then written out to a file as a standard wave file. This wave file is then forwarded to the Recognizer.

Recognizer

The Recognizer processes the file, returning the transcription of the utterance. Information is extracted using the Natural Language Interpretation. Please see the section on the speech recognition system for more information. The results are then sent to the Latent Semantic Analysis (LSA).

Control is then returned to the Transcriber. Here the audio, transcription, and results are logged. The results and transcription are displayed and color coded according to the confidence score from the speech recognition engine (Figure 5). The process ends by giving control back to the Courier.

JAVA DIS Tools

Several tools have been created that use the DIS library. These include a Scanner, Radio, and Logger tool. The Scanner tool, (Figure 6), allows live playback of audio traffic on the network. Just like a regular radio scanner, frequencies can be selected to play or be turned off. Each entity can also be turned on or off. The Scanner works much the same as the Transcriber, it separates each entity's PDUs. However, it doesn't buffer the entire utterance. It buffers a tenth of a second of audio and then starts the playback. More than one entity can be heard at a time. Active entities and frequencies are color coded with green representing currently transmitting entities and red representing online but not currently transmitting. The Radio tool (Figure 7), is build upon the Scanner. It has the same functionality as the Scanner, but adds the feature of talking live across a DIS network.

Figure 5. *Transcriber Tool*

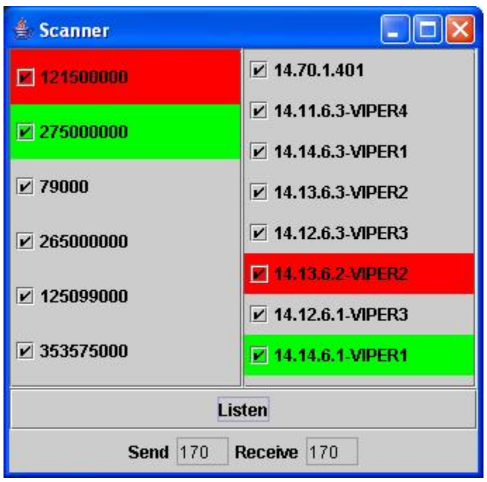


Figure 6. *DIS Scanner Tool*

Figure 7. *DIS Radio Tool*

Discussion

The tools described in this paper provide an initial capability to help automate the communications assessment process in a DMO environment. Initial performance results indicate that less than 20 percent word error rates (WERs) are achievable with the current statistical language modeling technique, which exceeds the initial program goal of 40 percent WER. Additional research is underway to attempt to improve upon this baseline performance. This includes producing separate language models, one for AWACS and one for the fighter group, as well as using additional data sets for the SLMs.

The use of separate SLMs would allow a more representative modeling of the specific communication events for each group. This would require multiple SLMs running simultaneously with the Recognizer module routing the specific command to the appropriate model based on the entity ID contained in the signal PDU.

The performance of SLM-based speech recognition systems is based on the data used in training the models. The use of additional data sets will result in better overall coverage of the domain and improve accuracy.

The more general purpose DIS Scanner and Radio tools are also being refined and expanded upon to increase their functionality. One enhancement to the Radio tool is the incorporation of 3D audio spatialization which combines location information from GPS and head tracker sources with the audio data to present localized audio to the players. Another enhancement is the ability to receive text messages over signal PDUs containing application specific data and converting them to speech for simulating additional communications traffic.

References

Air Force Instruction 11-2F-16, V1, 08 August 2002, Flying Operations, F-16 – Aircrew Training, Retrieved on May 10, 2004 from http://www.e-publishing.af.mil/pubfiles/af/11/afi11-2f-16v1/afi11-2f-16v1.pdf.

Air Force Policy Directive 33-1, Command, Control, Communications, and Computer (C4) Systems. 17 September 1993. Retrieved on May 10 2004 from http://www.e-publishing.af.mil/pubfiles/af/33/afpd33-1/afpd33-1.pdf

Department of Defense Joint Technical Architecture, Version 6.0 (2003). Retrieved on May 10, 2004 from http://jta.disa.mil/jta/jta-vol-I.pdf.

Institution of Electrical and Electronic Engineers, "Standard for Distributed Interactive Simulation – Application Protocols," IEEE Standard 1278.1a-1998, Mar. 1998

Laham, D., Bennett, W., & Derr, M. (2001). Latent Semantic Analysis for Career Field Analysis and Information Operations.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.

Williamson, D.T. & Snyder, R.A. Jr. (2002). Performance Assessment of a COTS Speech Recognition System on the N4 Database. (Tech. Rep. No. AFRL-HE-WP-TR-2002-0249). Wright-Patterson Air Force Base, OH: Human Effectiveness Directorate.

HEAD-UP DISPLAY SYMBOLOGY FOR SURFACE OPERATIONS: EYE TRACKING ANALYSIS OF COMMAND-GUIDANCE VS. SITUATION-GUIDANCE FORMATS

John Wilson[1], **Becky L. Hooey**[1], **David C. Foyle**[2]
[1] San Jose State University at NASA Ames Research Center, Moffett Field, CA 94035
[2] NASA Ames Research Center, Moffett Field, CA 94035

This study investigated pilots' taxi performance and distribution of visual attention with four different head-up display (HUD) symbology formats: Command-guidance, Situation-guidance, Hybrid, and a baseline, No-route guidance. Taxi speed and centerline accuracy were highest with Hybrid and Situation-guidance whereas Command-guidance and No-guidance resulted in increased visual attention to the head-down map display and side window displays. These results are thought to be due to lack of sufficient preview information with the Command-guidance symbology. The conformal route information of the Situation-guidance and Hybrid HUD formats provided a common reference with the environment, which may have supported better distribution of attention.

Introduction

Airport surface operations have been cited as the least technologically advanced and one of the most difficult phases of flight (Kelley & Adam, 1997). Pilots must maintain awareness of their cleared taxi route, their position relative to the cleared route, as well as their position on the airport surface. To maintain awareness, pilots must monitor airport signage and markings and compare this information to a paper airport diagram. In low visibility or at night, pilots often reduce their taxi speed to avoid traffic conflicts and maintain adequate position awareness.

One way that low-visibility surface operations may be improved is by using Head-Up Displays (HUDs) to depict the cleared taxi route (Foyle, et al., 1996). There are two general HUD symbology concepts for providing navigation information: Command-guidance and situation-guidance. Command-guidance symbology directly provides commanded control information and is commonly displayed as a non-conformal error from the ideal path. In contrast, situation-guidance symbology provides navigational information as a conformal representation of the path without displaying the required control inputs or the error deviation (Foyle et al., 2002).

Command-guidance symbology provides the pilot with information related to the control inputs required to minimize deviations from the cleared route. The pilot's role in such a system has been described as a "low-level servo" (Beringer, 1999). Examples of command-guidance symbologies are displays used in most current commercial aircraft that incorporate an aircraft reference symbol, flight director and command-guidance cue. In flight simulations, pilots flying with command-guidance HUDs fly with less error, both vertical and horizontal, compared to head-down command-guidance and head-up pathway symbologies (Weintraub & Ensing, 1992).

One potentially negative aspect of command-guidance symbology is that it produces more control inputs than other displays (Beringer, 1999). This is due to the command-guidance symbology constantly displaying guidance information as error from the ideal course, so that even small deviations require a course correction. This leads to the pilots making small s-turns about the ideal course. Also, it has been hypothesized that command-guidance symbology does not support efficient division of attention between the HUD symbology and the out-the-window environment (Foyle, et al., 1992, Foyle, McCann & Shelden, 1995), because it is often presented as a superimposed symbology at a fixed-location on the HUD. The resulting differential motion between the fixed-location symbology and the dynamic, out-the-window scene can lead to attentional fixation on the command-guidance symbology (McCann, Foyle & Johnston, 1993).

Situation guidance symbology presents the cleared taxi route by augmenting the environment with conformal, scene-linked symbology (Foyle, McCann & Shelden, 1995). It is conformal in that the symbology overlays and moves in unison with the environment (Ververs & Wickens, 1998) and it is scene-linked in that it represents objects placed in the actual environment with appropriate optical motion cues (Foyle, et al., 1992). Situation-guidance symbology does not provide the pilot with specific control inputs necessary to track the route, but instead augments the visual scene to allow the pilot to use external cues. A potential benefit of situation-guidance symbology is that it provides a better understanding of the desired path relative to current aircraft position and enables more effective path recovery as compared to command-guidance symbology (Beringer, 1999). Also, it has been shown to reduce cognitive tunneling, compared to fixed-location symbology (Foyle, McCann & Shelden, 1995). In sum, the benefits of situation-guidance symbology indicate improved attention distribution; however, this may come at a cost of increased tracking error (Beringer, 1999).

Previous Research

A previous study (Foyle, et al., 2002; Wilson, et al., 2002) was conducted to compare pilot performance using three different types of HUD symbology: Command-guidance, Situation-guidance and a Hybrid symbology that combined aspects of the Command-guidance and Situation-guidance displays. It was hypothesized that compared to the Command-guidance symbology, pilots taxiing with the Situation-guidance symbology would have higher taxi speeds and better situation awareness, but at the cost of increased centerline deviation. Since the Hybrid symbology combined elements from both formats, it was hypothesized that it would lead to increased taxi speeds and better situation awareness, with no subsequent increase in centerline deviation.

As hypothesized, when pilots taxied with the Situation-guidance and Hybrid symbologies, they had significantly higher taxi speeds compared to the Command-guidance symbology. It was hypothesized that centerline deviation would be least with the Hybrid and Command-guidance symbology, because of the command-guidance cue. However, results showed that while pilots had the least deviation with the Hybrid symbology, they actually had more deviation with the Command-guidance symbology compared to the Situation-guidance symbology. It was concluded that the increased centerline deviation with the Command-guidance symbology was due either to aspects inherent to the Command-guidance symbology concept, or to the specific symbology presentation that was instantiated in the study. Specifically, the Command-guidance symbology included a guidance cue and a graphical plan-view representation of the centerline. The plan-view centerline provided preview of approximately 100 ft. of the upcoming taxi route. This form of preview may have been insufficient as pilots referenced the head-down taxi navigation map for upcoming turn information. When the pilot went “head-down”, this may have contributed to the decreased taxing accuracy. To better understand this finding of decreased accuracy, the present study was conducted implementing three changes to the previous study. First, the Command-Guidance Symbology was modified to investigate the effect of preview in the form of an arrow with a text-based turn distance countdown instead of a plan view centerline. Second, the use of an eye tracker to record eye movement data was added to determine whether pilots’ distribution of visual attention differs as a function of symbology type and to address questions related to symbology usage. Third, a baseline condition was added to evaluate taxi performance and visual attention with each HUD symbology condition relative to current-day, no guidance, conditions.

Method

Participants

Fourteen commercial airline captains, thirteen male and one female, participated in the study. The pilots’ age ranged from 33 to 54 years (M= 44 yrs). The flight hours logged as captain ranged from 1,000 to 12,000 hours with a mean of 4,503 hours. All of the participants were certified by their airline to use a HUD, and HUD hours logged ranged from 250 to 8,000 hours (M=2,223 hrs).

Apparatus

A medium-fidelity part-task simulator at NASA Ames Research Center was used. The airport was Dallas-Fort Worth International Airport (DFW) with a visibility of 1200 ft runway visual range (RVR). The airport environment included terminal buildings, runways, taxiways, grass medians, taxiway signage and markings, moving and non-moving aircraft and ground vehicles. Aircraft controls included a side-stick tiller control with left/right rotation for nose-wheel steering, non-differential throttle and rudder pedals with toe brakes. The aircraft control model closely resembled a Boeing 737. Eye tracking data was collected using an Applied Science Laboratories (ASL) 5000 Integrated Eye/Head tracking system at a data collection rate of 60 Hz.

Out-the-window scene. The forward out-the-window scene was rear projected on a 2.44 m horizontal (H, 53.13 deg) by 1.83 m vertical (V, 41.11 deg) screen located 2.44 m in front of the pilot’s eye point. The HUD symbology was graphically presented on the forward screen, such that the HUD display area was 31.42 deg (H) by 15.60 deg (V). The side window scenes, subtending a visual angle of 29.57 deg, were presented on two 48.26 cm (19-in diagonal) monitors, one on each side, at a viewing distance of .91 m.

Map and clearance display. A north-up taxi chart of DFW was copied onto a transparency and overlaid on a computer monitor with a white background. At the bottom of the monitor was a text display with the taxi clearance for each trial. The map and clearance display area was 33.02 cm (H) by 24.13 cm (V) at a viewing distance of 1.07 m (17.54 x 12.87 deg).

HUD Symbology

Four HUD symbology formats were developed to explore performance and symbology usage differences among Command-guidance, Situation-guidance, Hybrid, and No-route guidance symbologies. All symbology types had text taxiway labels and a groundspeed indicator as shown.

The Command-guidance symbology (Figure 1) is composed of a command-guidance cue, turn-distance countdown, and turn-direction indicator. The guidance cue is similar to command-guidance symbology commonly used for maintaining flight path in the air (Weintraub & Ensing, 1992; Foyle, Hooey, Wilson, & Johnson, 2002). The inner circle, the command-guidance cue, moves left and right in relation to the outer circle (fixed aircraft reference symbol) based on taxiway centerline deviation. The pilot's task is to taxi the aircraft such that the two circles are concentric, which will result in recapturing or maintaining the centerline of the cleared taxi route. The turn-direction arrow and turn-distance countdown provided preview for the next turn in the cleared route.

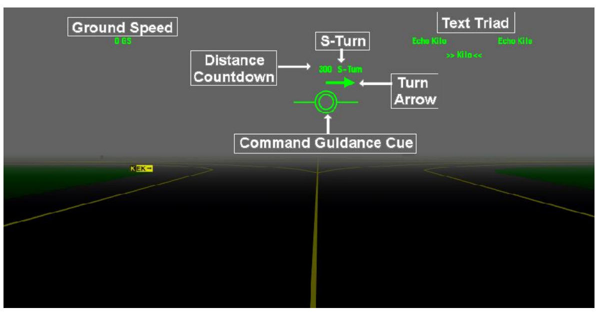


Figure 1. *Command-guidance symbology (symbology in green, white, boxed labels not shown in experiment)*

The Situation-guidance symbology (Figure 2) uses the HUD format of the Taxiway Navigation and Situation Awareness System (Hooey, Foyle & Andre, 2001). The cleared route is augmented with scene-linked symbology that overlays objects in the world including an augmented taxiway centerline, taxiway-edge cones, turn signs, and turn flags which extend beyond the cones in turns.

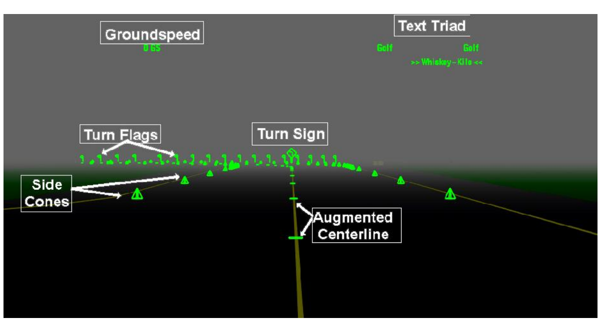


Figure 2. *Situation-guidance symbology*

The Hybrid symbology (Figure 3) combines aspects of the Command-guidance and Situation-guidance symbologies by providing control commands as well as conformally highlighting the cleared route. In the Hybrid symbology, there is a command-guidance cue, but without the turn-arrow or turn-distance countdown of the Command-guidance symbology. The Hybrid symbology has the scene linked taxiway edges and centerline of the Situation-guidance symbology without the turn flags and signs.

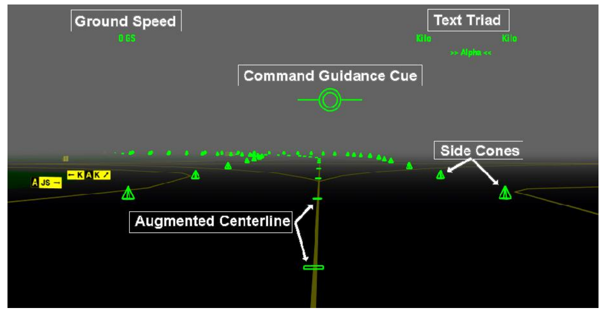


Figure 3. *Hybrid symbology*

The No-route guidance symbology (not shown) was implemented to simulate current taxi operations. The HUD provided only a ground speed indicator and taxiway text labels. This was provided in lieu of a first officer, who would normally assist the captain by calling out the upcoming taxiways.

Experimental Design

The study was a within-participants design, with HUD symbology format (Command-guidance, Situation-guidance, Hybrid and No-route Guidance) as a four-level factor. Each participant completed 16 experimental trials: Four consecutive trials of each of the four HUD formats. Order of presentation of the HUD symbology formats was randomized.

Scenarios. All scenarios consisted of taxi-only routes, with no landing or take-off. On average, the taxi routes were 15,600 ft in length, and contained six 90-degree turns. Each experimental taxi trial required approximately 8.75 minutes to complete, such that the entire experiment required a full day of testing. Each scenario included other aircraft and airport vehicles that were included for simulation realism and evaluation. (A near-incursion and situation awareness probes were included but are not reported here).

Procedure

Simulator training and familiarization consisted of eight trials (two trials each of the four HUD symbology formats) presented in randomized order. Through these training trials, pilots experienced instances of all scenario events with the exception of an aircraft incursion, and were briefed on the appropriate procedures for responding to the events. Upon completion of training, participants completed four blocks of four trials each, with a 10 minute break between each block. During each experimental trial, pilots followed a taxi clearance that was presented by voice from a pseudo air traffic controller (the experimenter) as well as presented in text on the map display. Pilots were told to taxi as they would in the real world with a full commercial flight and that taxi speed, accuracy, and safety were all equally important.

Results

Taxi Performance

Taxi performance with the four HUD symbology formats was assessed with two dependent variables: Average moving (non-zero) taxi speed (kts) and Root Mean Square Error (RMSE) from taxi centerline.

Taxi Speed. Increased taxi efficiency is one of the goals of a taxi HUD. Therefore, average moving taxi speed is an important measure of performance. It also serves as a surrogate measure of a pilot's confidence, as pilots taxi slower with greater navigation uncertainty. Consistent with the previous study, taxi speed differed as a function of HUD symbology, $F(3,39)=17.57$, $p< .001$. Taxi speed was greatest with the Situation-guidance (M=19.00 kts) and Hybrid (M=18.72 kts) symbologies, with no significant difference between the two. Situation-guidance and Hybrid symbologies were both significantly faster than the Command-guidance symbology (M=16.38; $t(13)=5.29$, $p<.001$, $t(13)=5.30$, $p<.001$, respectively) and the no-guidance symbology (M=16.26; $t(13)= 5.66$, $p<.001$; $t(13)=3.81$, $p<.01$, respectively). Presumably, the situation-guidance elements (enhanced centerline and cone augmentations) common to both the Situation-guidance and Hybrid symbologies better supported efficient taxi and navigation awareness than the command-guidance cue. Interestingly, Command-guidance did not yield increased taxi speeds over the No-guidance condition.

Taxi Accuracy. A second goal of the taxi HUD is to improve taxi accuracy, measured here as Root Mean Square Error (RMSE) deviation from the centerline. Recall that although Command-guidance was expected to produce superior taxi accuracy over the other symbology type, this hypothesis was not supported in the previous study. The current study aimed to further investigate this surprising finding with a different form of Command-guidance symbology (text vs. graphical turn preview). The results of the current study replicated the previous study. The RMSE data averaged across the entire trial, and averaged over turns only, are presented in Figure 4, however, as the results were identical, only the overall results are discussed. Taxi accuracy varied as a function of HUD symbology, $F(3,39)=11.94$, $p<.001$. The Command-guidance and No-guidance symbologies had the highest RMSE and were not significantly different. RMSE with Command-guidance symbology was higher than with Situation-guidance, $t(13)=2.30$, $p<.05$, and Hybrid, $t(13)=6.36$, $p<.001$. RMSE with No-guidance symbology was also higher than with Situation-guidance, $t(13)=3.70$, $p<.01$ and Hybrid, $t(13)=4.40$, $p<.001$. There was no significant difference between the Situation-guidance and Hybrid symbologies.

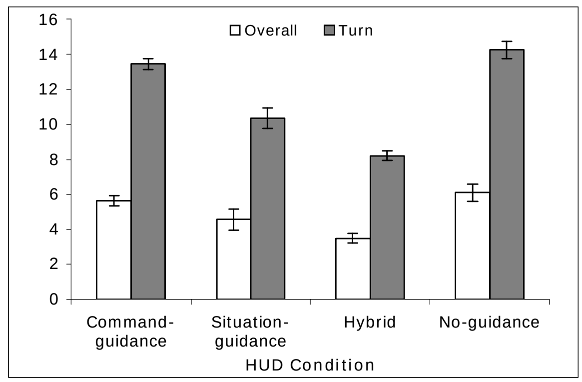


Figure 4. *RMSE averaged over entire trial and turns only, (+/- 1 SE).*

Pilot Visual Scanning

An important aspect of this study was to determine how pilots allocate visual attention while taxiing with different symbology types. There were three areas of interest: the taxi navigation map display; the forward screen, representing the forward aircraft window and HUD symbology; and the side monitors, representing the aircraft side windows. Figure 5 depicts the percent of total time allocated to each area over the entire trial for each HUD condition.

Allocation of Visual Attention to the Forward View. The forward screen provided a 53 deg (H) field of view of the airport taxiways and traffic immediately in front of the ownship and the HUD symbology overlay. Given the importance of scanning the environment for traffic and maintaining forward navigation awareness, it can be assumed that pilots only glanced away from the forward screen when they needed to gather navigation information from the map or side monitors that was not otherwise provided in the forward scene or symbology. Figure 5 demonstrates a significant difference in forward screen usage among HUD conditions, $F(3,39)=35.15$, $p<.001$. Pilots spent the most time looking at the forward screen when taxiing with the Hybrid symbology, which was significantly more than Situation-guidance, $t(13)=2.19$, $p<.05$. Both the Hybrid and Situation-Guidance displays yielded more time on the forward screen than either the Command-guidance, ($t(13)=4.25$, $p<.01$, $t(13)=2.28$, $p<.05$, respectively) or No-guidance, ($t(13)=9.09$, $p<.001$; t(13)=8.11, $p<.001$, respectively). Pilots allocated more time to the forward screen with Command-guidance symbology than No-guidance, $t(13)=7.81$. $p<.001$.

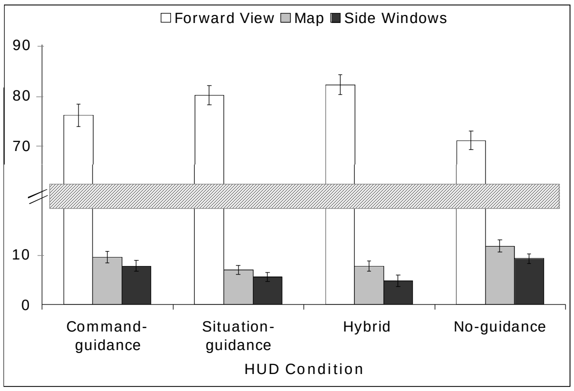


Figure 5. *Percent Time of Visual Attention Allocated to Forward View, Map, and Side Windows. (+/- 1 SE)*

Allocation of Visual Attention to the Taxi Map. The taxi map provided navigation awareness information about the cleared taxi route and the ownship distance to, and direction of, the next turn. The amount that pilots relied on the map for this information differed as a function of the HUD symbology, $F(3,39)=8.70$, $p<.001$. Specifically, when taxiing with the two symbologies that possessed situation-guidance elements (Situation-guidance and Hybrid), pilots spent less time viewing the map than when taxiing with the Command-guidance display, ($t(13)=2.45$, $p<.05$; $t(13)=2.97$, $p<.05$, respectively) or No-guidance ($t(13)=4.53$, $p<.001$; $t(13)=4.09$, $p<.001$, respectively. Situation-guidance and Hybrid formats yielded approximately equivalent map usage, as did Command-guidance and No-guidance. Presumably, the situation-guidance information available in the Situation-guidance and Hybrid symbologies provided more navigation awareness information than was available in the Command-guidance or No-guidance conditions.

Allocation of Visual Attention to the Side View. The side monitors were used by pilots for navigation (i.e., reading airport signage) and to follow centerlines through turn maneuvers. The total percent of time that pilots allocated their visual attention to the side monitors differed among HUD conditions and is shown in Figure 5, $F(3,39)=11.80$, $p<.001$. Overall, pilots spent the most time looking at the side monitors when taxiing with the No-guidance symbology, compared to when taxiing with the Command-guidance symbology, $t(13)=2.93$, $p<.05$, Hybrid symbology, $t(13)=4.2$, $p<.01$, and Situation-guidance symbology $t(13)=3.52$, $p<.01$. The Command-guidance symbology yielded significantly more time looking at the side monitors than did the Situation-guidance, $t(13)=2.60$, $p<.05$, and Hybrid, $t(13)=3.82$, $p<.01$, symbologies. Pilots spent the least amount of time looking at the side monitors with the Hybrid and Situation-guidance symbology, with no significant difference between the two conditions.

It is particularly relevant to examine side monitor usage for turn performance alone because it is during the turns when pilots look to the side for navigation guidance, if it is not available in the HUD. As expected, the percent time on the side monitors was significantly different as a function of HUD symbology, $F(3,39)=14.57$, $p<.001$. Pilots, when maneuvering turns, spent more time looking at the side monitors with the No-guidance symbology (M=3.5%), than with Situation-guidance (M=2.2%), $t(13)=3.37$, $p<.01$, Command-guidance (M=2.0%), $t(13)=3.61$, $p<.01$, and Hybrid (M=1.0%), $t(13)=4.71$, $p<.001$. This reflects the need for additional information to support the turn that was not available in the front screen or HUD. Pilots taxiing with Hybrid symbology spent significantly less time looking at the side monitor in turns than did pilots taxiing with the Situation-guidance, $t(13)=3.65$, $p<.01$, and Command-guidance, $t(13)=3.09$, $p<.01$. There was no significant difference between Situation-guidance and Command-guidance during turns.

Guidance Cue Usage. Recall that the RMSE data showed that centerline deviations were lower with the Hybrid symbology than the Command-guidance symbology. Given that both symbologies included the same guidance cue for centerline tracking, this difference was somewhat of a surprise in this and the previous study. Recall also that the two conditions differed in that the Command-guidance symbology used the guidance cue as a primary navigation source, while the Hybrid symbology utilizes the guidance cue in conjunction with situation-guidance symbology.

To better understand the RMSE difference, the percent of forward screen time that pilots dwelled on the guidance cue was examined for the Command-guidance and Hybrid conditions. When averaged across the entire trial, there was not a significant difference in the time spent looking at the guidance cue with the Command-guidance and the Hybrid symbologies. However, there was a significant difference during turns. When maneuvering turns, pilots spent more time on the guidance cue with the Hybrid symbology (M=3.9%) than when taxiing with the Command-guidance symbology (M=2.5%), $t(13) = 2.52$, $p < .05$. This suggests that the situation-guidance elements embedded in the hybrid symbology may have supported taxiing with the guidance cue and yielded more accurate taxi. Without this information, pilots were forced to take their eye off the forward screen and guidance cue and rely on the map and side monitors for navigation information to supplement the guidance cue resulting in greater centerline deviation.

Discussion

The Command-guidance and the No-guidance symbologies produced the highest RMSE deviation from the centerline and slowest taxi speeds, while the Situation-guidance and Hybrid symbologies produced the lowest RMSE and the fastest speeds. The eye-tracker results provided insight into pilots' usage of the displays, which may help to explain the RMSE deviation seen with the Command-guidance symbology. Pilots taxiing with the Command-guidance symbology spent less time looking at the forward screen, and more time looking at the taxi map and side monitors than did pilots taxiing with Situation-guidance and Hybrid symbologies. When taxiing with the Command-guidance symbology, pilots may not have had as much route knowledge, through preview, as with the Situation-guidance and Hybrid symbologies. Pilots with the Command-guidance symbology may have been forced to rely more on the map and airport signage, through the side windows, to confirm their positions and upcoming turns.

Pilots spent more time using the guidance cue during turns with the Hybrid symbology than the Command-guidance. With the Hybrid symbology, pilots attended more to the forward screen and the taxi task, utilizing both the guidance-cue and situation-guidance elements, without having to utilize the map and side monitors, thus improving taxi accuracy as evidenced by the least centerline deviation. This suggests that the Hybrid symbology may better support turns.

Pilots exhibit better taxi performance when they spend more time attending to the forward screen and less time looking at the map and side monitors to determine their position. In general, the guidance cue as a stand-alone navigation tool (Command-guidance), without the aid of scene-linked navigation aids does not seem to support accurate taxiing. When used in conjunction with scene-linked navigation (Hybrid), the guidance cue enabled more accurate taxi performance. However, questions remain about whether that benefit of improved accuracy outweighs the possible cognitive tunneling on the guidance cue, resulting in reduced division of visual attention.

References

Beringer, D. B. (1999). Flight command-guidance indicators or situation-guidance displays: My way or the highway? *Proceedings of the 10th International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University.

Foyle, D. C., Andre, A. D., McCann, R. S., Wenzel, E. M., Begault, D. R., & Battiste, V. (1996). Taxiway navigation and situation awareness (T-NASA) system: Problem, design philosophy, and description of an integrated display suite for low-visibility airport surface operations. *SAE Transactions: Journal of Aerospace*, *105*, 1411-1418.

Foyle, D.C., Hooey, B.L., Wilson, J.R. & Johnson, W.A. (2002). HUD symbology for surface operations: Command-guidance vs. situation-guidance formats. SAE Transactions: Journal of Aerospace, 111, 647-658..

Foyle, D. C., McCann, R. S., & Shelden, S. G. (1995). Attentional issues with superimposed symbology: Formats for scene-linked displays. In R. S. Jensen, & L. A. Rakovan (Eds.), *Proceedings of the 8th International Symposium on Aviation Psychology*, 98-103. Columbus, OH: Ohio State University.

Hooey, B. L., Foyle, D. C., and Andre, A. D (2001). The design of aircraft cockpit displays for taxi operations. In A. G. Gale (Ed.) *Vision in Vehicles IX*. Holland: Elsevier Science Publishers.

Kelley, D. R., & Adam, G. L. (1997). The human factors of runway incursions caused by "pilot error:" A survey of U.S. airline pilots. In R. S. Jensen & L. Rakovan (Eds.) *Proceedings of the 9th International Symposium on Aviation Psychology*, 911-917, Columbus, OH: Ohio State University.

McCann, R. S., Foyle, D. C., & Johnston, J. C. (1993). Attentional limitations with head-up displays. In R. S. Jensen(Ed.), *Proceedings of the 7th International Symposium on Aviation Psychology*, 70-75. Columbus OH: Ohio State University.

Ververs, P. M., & Wickens, C. D. (1998). *Conformal flight path symbology for head-up displays: Defining the distribution of visual attention in three-dimensional space*. University of Illinois Institute of Aviation Technical Report (ARL-98-5/NASA-98-1). Savoy, IL: Aviation Research Lab.

Weintraub, D. J., & Ensing, M. (1992). *Human Factors issues in head-up display design: The book of HUD* (SOAR 92-2). Dayton, OH: CSERIAC.

Wilson, J. R., Hooey, B. L., Foyle, D. C., & Williams, J. L. (2002). Comparing pilots' taxi performance, situation awareness and workload using command-guidance, situation-guidance and hybrid head-up display symbologies. *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomic Society*. Santa Monica, CA: HFES.

Acknowledgement

This research was funded by NASA's Airspace Operations Systems/HMP/HADMT program. We are grateful to Walt Johnson (Rockwell Collins Flight Dynamics) for developing and testing the command-guidance symbology used in this study.

WHY'D THE THEY DO THAT?
ANALYZING PILOT MINDSET IN ACCIDENTS AND INCIDENTS

Ed Wischmeyer, Ph.D., ATP/CFII
Embry-Riddle Aeronautical University
Prescott, AZ

In teaching a series of classes that analyzed "classical" airline accidents, it was observed that human behavior in such accidents was often understood better when the "mindset" of the protagonists was studied explicitly, rather than implicitly, and that the time element was also useful in such analyses. Often pilots took actions not explainable by traditional error models, and not predictable from known influences. These observations started a series of iterations that eventually converged on the PEEMBO model (Predispositions, Environment, Events, Mindset and mental condition, Behaviors, and missed Opportunities.) The PEEMBO model now appears mature enough for routine use in accident and other analyses, and presents insights not obvious in SHEL, Reason's, and derivative models.

Summary

In teaching an introductory graduate level class which addressed "classical" airline accidents, it became apparent that a direct, rather than indirect, focus on flight crew "mindset" provided an extremely useful analysis perspective complementary to existing analysis models. This observation co-evolved with the PEEMBO (pronounced pim-bo) model, whose elements are Predispositions, Environment, Events, Mindset, Behavior, and missed Opportunities. Initial analyses of both NTSB accident reports and ASRS anecdotal reports indicate high value to this approach and model.

The central hypothesis of this note is that there is a significant class of airline accidents where pilot mindset is the central, identifiable, and addressable contributor to the accident. An example of mindset is this classic quote:

> *"When anyone asks how I can best describe my experience in nearly 40 years at sea, I merely say, uneventful. Of course there have been winter gales, and storms and fog and the like, but in all my experience, I have never been in any accident of any sort worth speaking about. I never saw a wreck and never have been wrecked, nor was I ever in any predicament that threatened to end in disaster of any sort". Edward J. Smith, Captain, RMS Titanic.*

The PEEMBO model was evolved to support analysis of accidents that have a large human error component, such as airline accidents. Key attributes of such accidents include:

- The accident is caused by the conjunction of multiple factors.
- The accident sequence usually progresses over a non-instantaneous time frame, with strong correlation over time between many observed phenomena.
- One or, more usually, many of the observed phenomena are best described by human factors.
- Actions and decisions of the flight crew are driven by the mindset (motivations and beliefs) of the crew, and although that mindset is in general influenced by predispositions, in incidents and accidents, there is often a stochastic element.
- The failure to employ available "defenses"[1] often appears to reflect mindset. If the flight crew believes their goal is achievable, there is no reason to abandon that goal by employing a defense.

The advantages of the PEEMBO model include:

- Focus on influences on the protagonist (e.g., the flight crew), not on a "probable cause."
- Focus on the mindset of the protagonist. This mindset may not be considered an actual "cause," but the protagonist's mindset may reveal. motivations for a variety of decisions made, actions taken, and defenses bypassed.
- Recognition of the time element in such accidents, where the relationship between significant phenomena may be described by vehicle dynamics or logical consequences.[2]
- Recognition of individual events as significant in the progression of the accident. Events have a specific time or a short time range associated with them, as opposed to environment.
- Differentiation between those factors under the control of the protagonist and those not.
- Identifies a remediable, hazardous attitude.

This paper discusses PEEMBO and attempts to place it within the context of other human error models.

---

[1] Reason, James.
[2] Leveson has a good discussion on "chains of (time ordered) events."

Evolution of PEEMBO

Determination of "probable cause" has been a goal of accident investigation at least since the founding of the National Transportation Safety Board. Presumably, the focus on probable cause is to identify specific factors and thus to permit the initiation of remedial action. Thus, other analyses that similarly permit the initiation of remedial action will ultimately achieve the same goal, even without the focus on "probable cause."

Traditional accident analysis classes frequently use the SHEL(L) model and Reason's model as ways of identifying such remediable factors. However, the combination of these two was often observed to be weak in describing the time sequence of events in an accident chain; the immediate mindset of the protagonist; the occurrence of significant chance events; and the degree of contribution of various elements. Leveson's STAMP model provides another excellent means of analysis, but is weak in accommodating variations in human motivation.

Similarly, the *information-decision-action* model of Nagel ignores "the possibility of… inappropriate high level goals." [3] The PEEMBO model would express some its analogous concepts as *inputs-mindset-behaviors*, hypothesizing that the information processing included in the first stage of Nagel's model is very heavily influenced by the mindset of the crew. Indeed, Nagel quotes Monan (1986): "Pilots heard what they expected to hear, heard what they wanted to hear and frequently did not hear what they did not anticipate hearing – amendments to just-issued clearances." To rephrase this, "pilots remolded what they heard to be consistent with their mindset," or to expand this further, "pilots remold their perceptions, decisions, and actions to be consistent with their mindset."

Similar to the PEEMBO model, but not part of its evolutionary chain, is the 5-M model by T.P. Wright of Cornell University: originally man-machine-medium, mission and then management were added.

The basic PEEMBO model is shown in Figure 1, with only the major relationships between blocks shown.

Elements of the PEEMBO Model

Predispositions are those pre-existing, repeatable factors that shape the way that the protagonist thinks about and reacts to the operating environment and events. Predispositions share the characteristic that they will be repeatedly observable in the protagonist over a period of months and years, across multiple situations and events.

Safety training, such as windshear avoidance training, CRM training, and runway incursion awareness training, are well known examples of attempts to reshape a pilot's predispositions to:

- Suspect the presence of dangerous conditions,
- Detect and mitigate dangers,
- Employ defenses and not miss opportunities.

Predispositions include self-image, the way the protagonist wishes to be perceived by others,[4] training, experience, policy, procedures, skills, confidence, values, beliefs, personality style, techniques in executing procedures (whether formally sanctioned and taught or learned ad hoc), and interpersonal communication style.

Training and self-improvement are common techniques used to improve an individual's predispositions. However, in the course of a single event, an individual will have no control over his predispositions.

The FAA teaches the "five hazardous attitudes,"[5] and these are predispositions. However, these attitudes are often inadequate to explain pilot mindset.

Environment refers to the operating environment for this particular event, particularly as it affects the style of human operation and the probability of various events. Examples of environment include visibility, weather, runways in use, thunderstorm, traffic level, competence of various individuals, capabilities of the aircraft, capabilities of other hardware, schedule pressure, equipment installed on the ground or in the air, equipment not operational on the ground or in the air, and many other factors.

The protagonist will have little control over the environment. Usually, the one choice available is to leave the environment. For example, an airline crew may choose not to land at some airport, leave the environment of that airport, and go elsewhere. Less commonly, a protagonist may attempt to reshape the environment by requesting that additional equipment be activated, or by requesting that another individual improve their performance (e.g., to "shape up!")

[3] Wiener and Nagel, Chapter 9

[4] Associate Dean Mike Polay, Embry-Riddle Aeronautical University, Prescott, AZ, personal conversation

[5] FAA Advisory Circular 60-22

The environment can influence mindset. Recognition of a hazardous environment can induce a cautious mindset, or schedule pressures can induce a risk-tolerant mindset. Similarly, a high workload environment can induce a mindset of excessive focus on the task at hand, with errors being a result. Similarly, such a high workload environment can increase the odds that opportunities are missed.

Events are occurrences external to the protagonist, beyond his control, that occur during the time frame of the event. Events can be considered in two ways:

- Events would commonly be thought to include lightning strikes, wind shear, mechanical failures, radio transmissions, an aircraft on the runway, and other chance events that occur within the environment being considered.
- Events are also changes in the environment. For example, a change in which runway is active is appropriately considered an event.

The likelihood of events is predictable from the environment, but not the occurrence of a specific event at a specific time. For example, the presence of a severe thunderstorm will predict lightning, rain, gusty winds, and wind shear, but will not determine when each lightning stroke will occur. Similarly, an environment of high radio traffic will increase the likelihood of a transmission being blocked.

Note that only the likelihood of events can be predicted, not a particular occurrence. For example, gusty winds can be predicted, but not the specifics of any one gust. Thus, events provide a way for "chance events" to be incorporated into the error model. Protagonists can only react to events, not control them.

A point implied by the PEEMBO model is that the protagonist will be susceptible to these chance events. The protagonist will be prepared for these chance events by training, equipment, experience, and perhaps warning systems, but those preparations frequently will be designed to reduce susceptibility to chance events, not to handle nor to avoid all chance events.

"External" implies that the PEEMBO model is contextual, and considers the context of the protagonist(s) as "the" context.

Mindset / Mental Condition In many accidents, the mindset of the protagonist is the central factor in the accident, and the decisions and actions taken, and the opportunities missed, are manifestations of this mindset. Thus, Mindset is central to the PEEMBO model.

Mindset includes the attitudes, motivations, expectations, knowledge, feelings, plans, goals, and self-image of the protagonist in this one situation. Examples of mindset are, "I can do this," "I have to do this," or "I'd better not do anything until I figure this all out."

An alternative expansion of the "M" in PEEMBO is "mental condition." The protagonist's mental condition can be strongly influenced by factors such as the perceived or anticipated degree of difficulty of the flight; complacency or apprehension; anticipated competency of others; impatience; expected environment; physiological stresses of all sorts; psychological stresses of all sorts; and knowledge of conditions, including observations of other flights & environment. Similarly, skills and competence are components of mental condition.

"Mental condition" seems a more complete, more clear, and more useful analysis tool than "mindset." However, many accident scenarios indicate that pilot mindset, unfortunately adopted and never challenged, is a major factor in many accidents.

Although predispositions are presumed to be relatively consistent over a period of weeks or months, the mindset of the protagonist is considered only within the context of each accident occurrence. Thus, mindset will not necessarily be consistent across situations.

Just as the protagonist will have no control over predispositions and events, and at most limited control over the environment, the protagonist may have little if any control over his mindset.

Explicit examples of mindset are not hard to find. In the Cali accident, at time 2136:38, the First Officer expressed mindset by saying, "We can do it." Similarly, in the Burbank accident, there is a strong implied mindset, "we can do this." Although this is speculative, one wonders at what point "we can do this" transforms into the dangerous mindset, "we have to do this."

In a set of 10 NASA ASRS reports on unstable approaches,[6] mindset statements included:

1. "We'll be all right." (ASRS 458452)
2. "Enough is enough." (ASRS 450568)
3. "I do not like the looks of this" and "Let's see how it is at 1000 feet" (ASRS 144766)
4. "[I said] we could make it." Later, the captain did in fact go around, indicating a more safety conscious mindset of, "we can make it, if…" (ASRS 253786)

[6] Selected by graduate student Mukul Mishra for a class on flight safety

5. “[he said] he could still do it.” (ASRS 302878)
6. “Try to save an approach visually.” (ASRS 521341)
7. “He tended to want to ‘fly solo.’” (ASRS 305526)
8.

Although no detailed study of mindset in ASRS approaches has been conducted, it is impressive to see such clear statements of mindset in such a high proportion of reports.

Leveson states “explanations for human goals and motives will depend on assumptions that cannot be *directly* [emphasis hers] measured or observed by the accident investigator.” ASRS reports indicate that mindset information is available, if not in accidents per se, at least in anecdotal voluntary submissions.

Personal conversations with pilots from an airline with a strong FOQA program conveyed another mindset. One pilot told me that when an approach may be exceeding allowable limits, he asks himself if completing the approach will be worth having to make explanations to the FOQA monitoring committee on why he did what he did.

One valuable source of mindset is the phrasing and tone of speech. Such phrasing and tone is not available in printed CVR transcripts, and the unavailability of such information is not consistent with national safety goals. There seem to be multiple ways of providing such valuable phrasing and tone information while meeting traditional privacy goals.

Behavior refers to both the decisions and actions of the protagonist.

During the progression of a scenario, the protagonist may make a number of decisions. These decisions may result in deliberate actions at either an abstract level (e.g., starting an approach) or at a lower level (e.g., calling for the landing gear to be lowered.) A decision, once made, will limit possible future decisions and actions. Once an action is taken, it, too, will limit future actions and decisions. For these reasons, actions and decisions are grouped into “Behaviors.”

A property of “behaviors” in the PEEMBO sense is that a time, or a start time before a relatively short time interval, is associated with each behavior.

Not all actions will be the result of conscious decision, however. For example, a skilled pilot may perform tracking tasks without conscious deliberation of each correction.

Reason’s error model of slip, lapse, violation, and mistake is split in the PEEMBO model: slip, violation, and mistake are considered behaviors, and a lapse is considered a missed opportunity, described below.

Missed opportunities are actions that the protagonist could have been taken to reduce risk or severity but did not. Missed opportunities also include decisions that were not made, said decisions establishing the framework for the actions not taken. Missed opportunities in the context of incidents and accidents may be defenses not employed in Reason’s model.

Examples of missed opportunities include going around; proceeding to the alternate; being more clear in communications; and asking for clarification of communications from others.

It seems clear that a flight crew will not employ defenses when they believe that the flight can be safely concluded. An interesting hypothesis is that the crew’s mindset causes them to ignore cues and misinterpret events, as hinted at by Monan (1986), above.

Missed opportunities are most interesting when they could have prevented the accident, but missed opportunities may be mitigating as well. Missed opportunities may or may not be associated with a specific point in time, and may or may not be repeatable. Useful classifications of missed opportunities include:

- Preventing – taking this course of action or making this decision would have prevented the accident
- Supporting – although this course of action not taken would not have prevented the accident, it supports observations about the mindset of the flight crew.

Causality

In our society, we tend to look for direct causes of accidents. Certainly this phenomenon is observable in the press, and sometimes in politicians, legal proceedings, and the military. Indeed, the idea that retribution will re-establish the world order is traceable at least as far back as Shakespearean times.

Particularly in aviation accidents, such direct causality is rare. More common, and more difficult to analyze are events in which improper motivation(s) were the “direct cause.” For such psychological phenomena, it seems more appropriate to use terms that reflect such motivations than to use terms of formal logic.

In the PEEMBO model, the following non-exclusive terms are used to describe the causality of a phenomenon contributing to an accident:

- Necessary – if this one phenomenon were not present, the accident would not have occurred, even if all other phenomena were still present.

- Sufficient – this phenomenon would cause the accident all by itself, even if all other factors were absent. Sufficient phenomena are rare in aviation, with the most common occurrences being irrecoverable mechanical problems.
- Continuing – this phenomenon is a direct consequence of a previous behavior (action or decision). Because it is a consequence, it adds little if anything to the discussion of "causality." For example, high airspeed on final approach following a late descent is not considered an independent phenomenon, but is considered to be a continuation of the late descent.
- Irrelevant – this phenomenon has nothing to do with the causality of the accident. It is worth noting that phenomena may be irrelevant because they had nothing to do with the accident, or because the model has no way to handle that phenomenon. Most accident analysis models will not be able to properly accommodate all phenomena of an accident, so within the context of each accident analysis model, some factors will be irrelevant.
- Motivating – explains why the crew made the choices they did. Such motivations would include corporate culture, schedule pressure, and how each crew member wanted to be perceived by others.
- Contributing – while this phenomenon was neither necessary nor sufficient, it contributed to the evolution of the scenario being studied, particularly to the severity of the scenario or to the difficulty in returning to a more normal state of operation. Speculation is that many contributing factors may be motivating factors.
- Supporting – gives useful insight into the environment or the crew's mindset, although it had nothing to do with the accident causality in a strict sense. Thus, a supporting factor cannot be necessary. For example, in the case of a runway overrun accident, an ignored GPWS warning on short final might provide insights on the mindset and mental condition of the crew but, strictly speaking, would be irrelevant to a taxi error made once off the runway. (This is similar to the epidemiological concept of "confounding.")

Temporal Analysis

The PEEMBO model not only groups factors by control and influence, but by time. There are three time spans shared between the six boxes:

- Observed consistently before and during the scenario – Predispositions only;
- Observed during the course of the scenario, but may change at specific times – Mindset / mental condition, and Environment;
- Observed at specific points in time during the course of the scenario – Events, Behaviors, and Missed Opportunities.

These time spans help clarify in which boxes specific observed phenomena should be grouped.

Immediate Safety Lessons

The most common mindset observed using the Peembo model, but certainly not the only mindset observed, is "we can do it." Some accident reports indicate that this mindset is present in pilots whose personality does not fit the FAA "hazardous attitude" of "macho."

This suggests, then, that a simple safety improvement is that "we can do this" should never be accepted, but should always be conditioned. Examples are "we can do this if," "we can do this while," "we can do this until," and "we can do this unless…"

Conclusions

A direct focus on pilot mindset, as embodied in the PEEMBO model, has been shown in the classroom to be a valuable tool for analyzing airline accidents and events. A preliminary review of ASRS events suggests that mindset is an identifiable, addressable, and significant contributor to unstable approach events.

Greater application of this model to a broader base of scenarios will lead to improvements in this model and even better models in years to come.

Acknowledgements

The PEEMBO model was developed to support a Special Projects course in Flight Safety taught in the Master of Science in Safety Science program at Embry-Riddle Aeronautical University, Prescott, AZ. Student feedback, especially from the second year, has greatly improved this work.

References

Bruseberg, A and Johnson, P (2003) "Understanding human error in context: approaches to support interaction design using air accident reports." In: Jensen R (ed.): *Proceedings of the 12th International Symposium on Aviation Psychology*, 14-17 April 2003, Dayton, Ohio, USA, pp 166-171. (http://www.cs.bath.ac.uk/~anneb/humanErr.pdf)

___, Advisory Circular 60-22, Federal Aviation Adminstration, Washington, D.C.

Leveson, N. (2001). *Evaluating Accident Models Using Recent Aerospace Accidents*: NASA Internal Study, avail. http://sunnyday.mit.edu/accidents.

Monan, W. P. (1986) *Addressee problems in ATC communications: The call sign problem.* NASA Contractor Report 166462. Moffett Field, CA: NASA-Ames Research Center.

Reason, James. Human Error. Cambridge University Press, 1990.

Tillyard, Eustace M. Elizabethan World Picture, Vintage/Ebury (A Division of Random House Group), 1975

Wiener, Earl L., and Nagel, David C. Human Factors in Aviation, Academic Press, Inc. 1988. San Diego, etc.

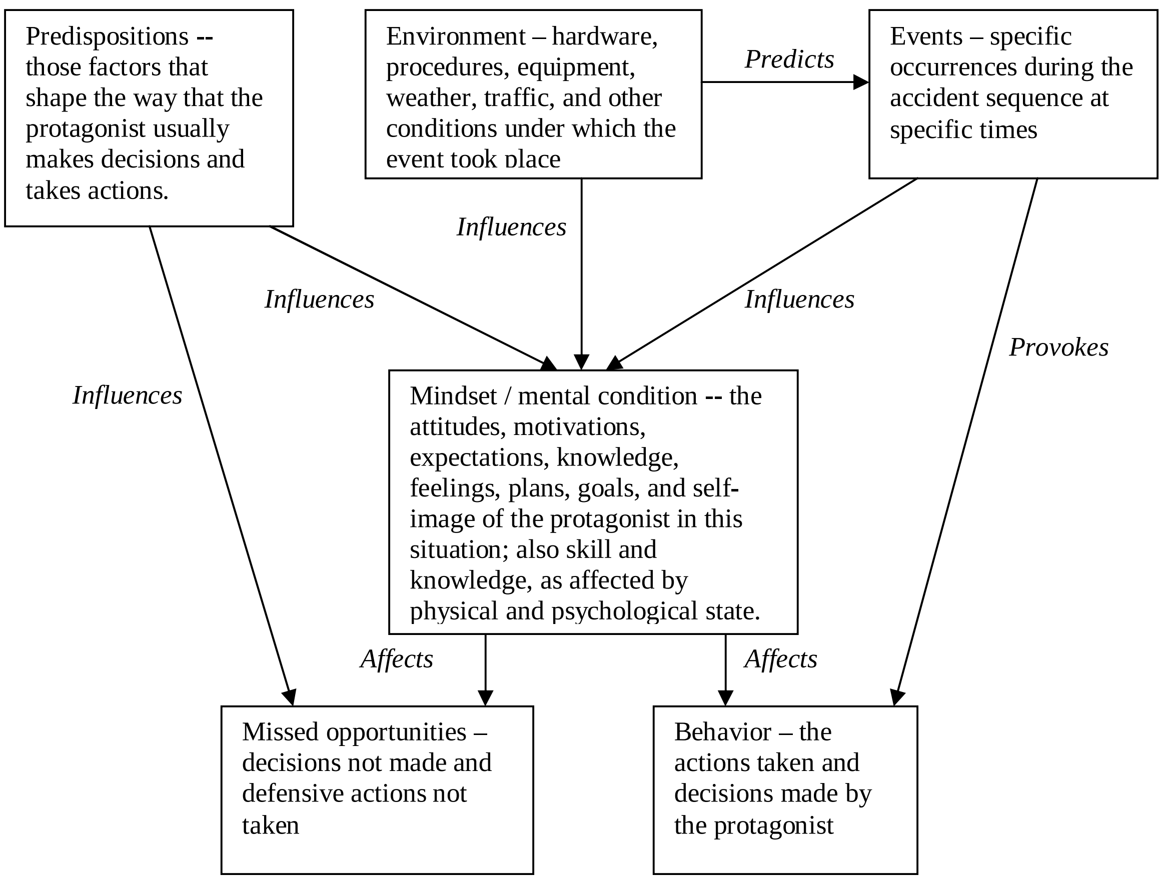


Figure 1. *Elements of the PEEMBO model*

METRICS OF INFORMATION COMPLEXITY FOR AIR TRAFFIC CONTROL DISPLAYS

Jing Xing
FAA Civil Aerospace Medical Institute, Oklahoma City, OK 73125
Jing.xing@faa.gov

Information complexity associated with visual displays is a bottleneck that limits their use. While automation tools are designed to bring new functions to users and increase their capacities, they also creates new tasks associated with acquiring and integrating information from displays. In particular, a complex display increases information load to human operators and reduces usability. Thus the efficiency of the tool largely depends on the complexity of displayed information. To evaluate the costs and benefits of an automation system, it is important to understand how much information is shown on the display, and whether the information is too complex for users to process. In this paper, we present a set of observable metrics to assess information complexity of visual displays. The metrics count information complexity as the combination of three basic factors: numeric size, variety, and relation; each factor is evaluated by the functions at three stages of brain information processing: perception, cognition, and action. Ideally, these measures provide an objective method to evaluate automation systems for acquisition and design prototypes.

Introduction

Many automation tools are developed provide decision-support information for air traffic control (ATC). While these tools are intended to ensure safety and offload tasks from controllers, they also create new tasks associated with interface management. In particular, information provided by the tools can be too complex and overwhelm controllers' cognitive capacities. Consequently, key information could be either missed or misinterpreted by controllers and thereby increases the risk of performance errors. For these reasons, it is desirable to have an objective method to assess information complexity (IC) of automation displays and to assess the impact of complexity on operators' task performance.

Most previous human factors studies have focused on how information should be presented, not necessarily information complexity (Tullis 1985; Sears 1994), although the latter has been theoretically explored. Information theories consider a system as an automaton consisting of a series of elementary units distributed in space. From the viewpoint of information theories, the most straightforward definition of IC is the minimum description size of a system (Grassberger 1991; Crutchfield & Young, 1989). That is to say, if the description of a system can be greatly compressed without loss of meaning, then it is considered simpler than one that cannot. However, this definition is only concerned with the storage demands of a system. In contrast, Bennett (1990) introduced the concept of logical depth as a measure of complexity. Logical depth combines resource demands and computational power into a single description of the computational resource required to calculate the results of a program of minimal length. This definition is a combination of both resource demands and computational power. Scott (1969), on the other hand, proposed a measure of information redundancy to describe complexity. Similarly, Langton (1991) suggested that complexity is associated with high levels of mutual information, which is the correlation between information at separated sites. In general, these studies focus on the difficulty of compressing a representation, with little direct connection to the practical aspects of a functioning organism. In addition, information theories define information in relation to the probabilities of all other inputs that might have been encountered. However, it is difficult to specify probabilities when applying theories like these to such realistic circumstances as ATC.

The objective of this report was to develop observable metrics of IC for automation displays. This objective raises three basic questions: What is complexity? Why can information be too complex for the human brain? Finally, how do we quantify the complexity of visual displays? We address these questions in this report by presenting a set of IC metrics developed for automation displays in ATC.

Results

Information complexity

Xing and Manning (2005) generalized the following definition of information complexity in visual display: Complexity consists of three basic factors: numeric size, variety, and relation; these factors are evaluated by users' mechanisms of information processing, and they are constrained by task requirements.

Given that complexity depends on how observers process information, we looked into the mechanisms of information processing in the human brain. Figure 1 outlines a conceptual diagram of human visual information processing associated with the use of visual displays. In this simple representation, information presented via visual display devices is processed by three stages in the human brain: perception, cognition, and action. Through perception, a user acquires information about the current status of the world. The perceived information then feeds into the cognition stage, where one's perceptions are integrated with information from the observer's experience and memory. An internal (mental) representation of what was observed can then be generated. Based on this representation and personal strategies, the observer can then make decisions and convert them into actions. The actions allow interaction between the observer and the system. These three stages have distinctive neural mechanisms and serve different brain functions, as briefly described below:

Perception – The human visual cortex is specialized to perform many kinds of perceptual functions including target searching, text reading, color discrimination, texture segmentation, motion detection, and many others. Perception processes information serially and in parallel. Thousands of visual neurons first extract information rapidly in parallel, the visual system then serially focuses the fovea on salient spots so that information can be analyzed in detail.

Cognition – The high-level modules of brain cortical areas, called associational cortex, integrate inputs from the perceptual cortex with information stored in brain's long-term memory. The associational cortex performs cognitive functions such as working memory, text comprehension, planning, selecting, etc. A common feature of cognitive functions is their limited capacity. That is, only a few pieces of information can be processed simultaneously in the associational cortex. Consequently, the bandwidth of information processing in the cognition stage is much less than that in the perceptual stage.

Action – The premotor and motor cortex of the brain are responsible for encoding various manual actions such as eye, head, hand, and arm movements. Those brain areas are also able to encode sequential movements. The motor cortex, unlike other cognitive and perceptual areas of the brain, is believed to work in a serial manner, i.e., all the neurons in the motor cortical area work together to encode a single movement and only after the movement command is executed do they begin to encode the next movement. Consequently, with such a narrow bandwidth of information processing, an effective automation tool should impose only very limited action requirements for human operators.

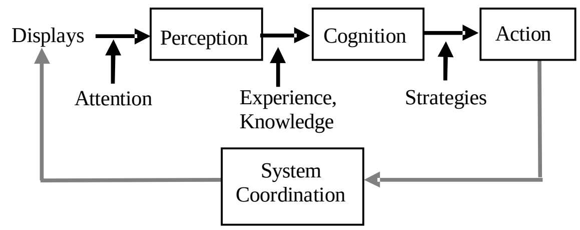


Figure 1. *A diagram of information processing in the human brain*

Given the differences inherent among the three information-processing stages, the three complexity factors should be evaluated separately at each stage. This results in a 3x3 matrix as shown in Table 1, with rows being the three complexity factors and columns being the three information-processing stages. Each box in the matrix corresponds to one IC metric. For each box, the complexity factor should be evaluated by the functions that occur at that stage, and each complexity metric should be associated with the operator's task performance. With the known capacity limits of brain functions, the metrics can elucidate why a visual display can be too complex for human operators. Table 1 lists the metrics we have proposed to assess complexity associated with automation displays in ATC. Each metric will be explained in the following sections.

Table 1. *Metrics of information complexity for ATC displays*

	Perception	Cognition	Action
Numeric Size	No. of fixation groups	No. of functional units (tasks)	Amount of keystrokes and mouse movements
Variety	Variety of groups	Dynamic change in the units	No. of transitions
Relation	Degree of clutter (Text readability)	No. of variables in a unit	Action depth (steps) of a functional unit

Steps for Developing Complexity Metrics

If we evaluate every complexity factor against every function of the three stages of brain information processing, the resulting metrics would have too many dimensions. Fortunately, complexity metrics are constrained by task requirements. Only the functions critical to the given tasks are relevant. The following steps were used to develop the metrics:

1) Identify task requirements;
2) Determine corresponding brain functions pertinent to the task requirements;
3) Choose the metric that can reflect the impact of the complexity factor on the brain functions.

Task Requirements of Using ATC Displays

ATC systems have unique features that differentiate them from other applications. Below are some typical characteristics of ATC automation displays:

1) Displays contain mainly text, icons, and other binary graphical patterns (symbol, charts, etc.). Spatially continuous digital images are very rare;
2) Controllers look for particular information on displays to assist in decision-making;
3) Displays are dynamic: Information is continuously updated with the evolution of the traffic situation;
4) Unlike most human-computer-interaction systems, many ATC automation tools are presented as aids, not the objects that controllers have to operate on. Controllers use aids only when they are helpful (i.e., the benefit is greater than the cost) and ignore them when they are not.

Given these characteristics, we derived some basic requirements for using ATC automation. ATC displays must allow: 1) searching for information in a timely manner; 2) reading text reliably; 3) facilitating rather than disturbing decision-making; and 4) minimizing time-costing actions. The complexity metrics described below were developed to measure these requirements. For each metric, we will first introduce its definition and how it relates to ATC task requirements. We will then describe its impact on ATC performance and the capacity limit to address the question of "why information can be too complex for users."

Metrics of Information Complexity for ATC Displays

Metric-1: perceptibility

Size factor evaluated by perception

The proposed metric is the number of fixation groups. The basic element for searching and reading tasks is eye fixation. A fixation group is defined as a set of visual stimuli that can be grabbed with one eye fixation. Typically, a foveal fixation spans a view angle of about 2-4 degrees. The average time to search for a particular target on a visual display increases with the number of fixation groups. While there is no physiological limit on how many fixations one can make on a display, visual experiments have demonstrated that it takes 600-700ms for an observer to perceive the information in one fixation (Joseph, Chun, & Nakayama, 1997). Therefore, the capacity limit of this metric is determined by the time that a user has available to spend on an automation aid. For example, if a controller has 5s maximally to acquire the information from an automation aid, then the number of fixation groups included in the display should be less than 14 (5000/700).

In many applications, displays are very busy and it takes many fixations to view all the information. One strategy to reduce perceptual complexity is the use of color-coding, because information can be segregated into several categories with color-coding. Consequently, visual searching can be limited to the visual targets illustrated with a particular color. By doing so the number of fixations can be greatly reduced.

Variety factor evaluated by perception

This proposed metric is the variety of fixation groups. Variety is defined as the differences in visual features such as size, texture, luminance, contrast, and colors of the groups. Increasing the variety of visual features increases complexity. Visual studies have found that switching between visual features such as color and luminance contrast increases searching time. This effect is called "cost of switching." In addition, switches may also reduce the reliability of reading text and increase visual fatigue. Consider, for example, two figures (A and B) that contain the same text. The text in figure A has the same format while the text in figure B is manipulated in font, letter size, and luminance contrast to increase visual variety. As a result, Figure B will appear to be more complex than figure A due to its increased variety.

Relation factor evaluated by perception

The proposed metric is the degree of clutter. Clutter is the effect of masking the visual perception of a stimulus with the presence of other stimuli. Consequently, clutter can increase search time and reduce text readability. The effect is apparent when background visual stimuli are spatially superimposed on the text. Moreover, the perceived contrast of a visual target can also be largely suppressed by the presence of neighboring stimuli. The reduced luminance contrast results in deterioration of text readability and a corresponding increase in search time. Xing and Heeger (2001) examined this effect in a series of experiments. They found that the perceived contrast of a sine-grating patch embedded in a large patch of the same kind of gratings was about half the contrast perceived when the central patch was presented alone. However, when a blank gap was introduced between the central and surrounding patch, the suppression effect became much weaker. These experimental results implicitly suggest two methods that reduce the clutter effect: 1) reducing the amount of text in a display and, 2) reducing the continuity of graphics so that targets do not have immediate surrounds.

Metrics-2: Cognitive capacity

Air traffic control is cognitively demanding. Basic ATC tasks include monitoring, controlling, checking, diagnosing and decision-making. Many cognitive models of ATC have been proposed (Kallus, Barbarino, & Van Damme, 1997). The kernel of those models contains two components: mental representation (or "mental model" as it is called by some in the literature) and memory. Cognitive processing is based on a mental representation of the task environment. Mental representations of a given situation are built by organizing information into many independent entities that are kept on-line for awareness. On the other hand, working memory enables us to hold in our mind's eye the content of our conscious awareness, even in the absence of sensory inputs. In a sense then, working memory manipulates entities in one's mental representation. It links pieces of information that are simultaneously required for a particular task. Therefore, measures of cognitive complexity should quantify how much a task imposes demands on both mental representations and working memory.

Size factor evaluated by cognition

Given that a mental representation is the platform for cognitive processes, the size factor corresponds to the number of basic, independent elements in a given mental representation. The challenge is to define these entities with respect to the use of automation displays. These elements represent the essential characteristics of information provided by a display. A common strategy used to support cognitive processing is categorizing pieces of information, where categories represent independent dimensions that an operator comprehends. In this way categories correspond to the entities of a mental representation. It makes intuitive sense, then, that complexity would be greater when an operator views a display as having many categories and must make fine distinctions among those categories.

While categorization can be based on perceptual features, a number of studies have demonstrated that the categorization process in ATC task performance is mostly goal-oriented. "Goal-oriented" refers to any feature that is an important objective of the task. Therefore, the basic elements of a mental representation can be specified as the fundamental functional units of a display. Each of the units represents a distinctive objective of the tasks. The units are independent of each other and cannot be combined to a chunk. Hence, we defined the number of functional units as the metric of size factor evaluated by cognition. A display may have many functional units; each unit achieves specific functional goals. To use the display fully a user stores the functional units in the mental representation of the situation. Complexity therefore would logically increase with the number of units in a given display. As the number gets larger, the memory load could impair task performance; the user may either misinterpret the information or choose to ignore it. Conway and Engle (1996) reported that normal adults could actively maintain 9-16 independent items in their memory during the operation of a task. This limit is potentially related to the capacity of a mental representation.

Variety evaluated by cognition

The proposed metric can be specified as dynamic complexity, measured as the rate of information change over time. Information changes in a display impose cognitive loads in several ways: 1) increasing working memory load. Psychophysical experiments have demonstrated that a sudden onset of visual targets or even changes in luminance of visual patterns automatically takes working memory (Schmidt, Vogel, Woodman, & Luck, 2002); 2) reducing the stability of mental representation. To build a mental representation takes time. For example, it takes several minutes for air traffic controllers to "warm up" before their visual scan patterns become regular and they can reliably perform their tasks (Stein, 1992). As a result, if too many entities are updated at a high rate, the mental representation tends to deteriorate. That corresponds to controllers "losing the picture" (Hopkin, 1995).

Relation evaluated by cognition
A task can become more complex as the number of interacting factors increase. Thus complexity can be measured by the dimensionality of the relation or number of variables that are related in a task. We used the definition of relational complexity proposed by Halford et al. as the metric to describe how the relation factor of complexity affects cognition (Halford, Wilson, & Phillips, 1998). Relational complexity is defined as the number of independent elements or variables that must be simultaneously considered to solve a problem. Many cognitive processes, such as selection of actions, manipulation of goal hierarchies, reasoning, and planning actions, are examples of processing at high levels of relational complexity. Halford et al. argued that relational complexity reflects the cognitive resources required to perform a task. The more interacting variables that have to be processed in parallel, the higher both the cognitive demand and computational cost will be. For example, an equation $a = 3 * b$ is a binary relation while an equation $a/b = c/d$ is a quaternary relation and therefore more complex. Hence, relational complexity is suitable to measure the affect of relation on cognitive load.

Because working memory links pieces of information that are needed simultaneously for task performance, relational complexity turns out to be a straightforward measure of the working memory load of a task. Halford et al. further demonstrated that the processing capacity of working memory for normal adults is limited to quaternary relations: Adults can reliably integrate up to four relations in parallel while children can only integrate one or two relations. This quaternary limitation appears to be consistent with other studies that demonstrated the capacity limit of working memory at about four items (Cowan, 2001).

Metric-3: Action feasibility
The purpose of an ATC automation aid is to increase capability and decrease workload. Therefore, it is desirable that a display provides information without demanding too much action from users. This is especially important for time-critical tasks such as air traffic control. If an automation aid requires too many inputs from controllers, it shunts controllers' attention away from the main tasks and may increase the risk of operational errors. However, given that today's automation systems are designed to provide large volumes of information, they inevitably require controllers to interact with them. Specific actions may include 1) eye/head movements to search for specific information; 2) keystrokes to update information and make inquires and 3) mouse movements to select specific information on a display. The following metrics of action complexity were determined by quantifying how feasible it is to perform those movements in a timely manner.

Size evaluated by action
The proposed metric is the number of keystrokes and mouse movements. Compared with keystrokes and mouse movements, the time needed for eye and head movements is negligible. Therefore, only keystrokes and mouse movements are considered here. Mouse movements are typically made to select information in a region of interest (ROI). That is, the larger the area of an ROI, the less time that is needed to perform a selection action with the mouse. Thus the moving distance and the ROI size both contribute to the cost of mouse movements.

Variety evaluated by action
The proposed metric is the number of action transitions required by a functional unit. An action transition is a change of action modes, such as from keystrokes to mouse movements or vice versa. Those transitions take time and require the brain to coordinate different action modes.

Relation evaluated by action
The proposed metric is the degree of action depth needed to achieve the goal specified by a functional unit. Action depth is the number of serial steps needed to achieve the task goal of a functional unit. An example of action depth is the number of layers of pop-up windows needed to accomplish a given task. Complex systems are usually characterized by a multi-level structure. Theoretically, a two-level structure is desirable to maintain low complexity. With a two-level structure, the information hierarchy required by task goals is achieved by a number of parallel, independent subgoals. However, following the need to increase the variety of actions, today's automation aids tend to use multi-level structures to cope with more diverse environmental perturbations and reduce the difficulty of decision-making. In such systems, a task of any complexity can be decomposed into a series of subtasks each represented by a

subgoal. The subgoals are determined by interactions between the sub-structure of the original task and the details of the system interface. Researchers have used the number of serializable subgoals as a measure of complexity for a system with a multi-level structure (Heylighen, 1998).

Discussion

The metrics presented in this report are based on theoretical studies and field observation of ATC automation displays. They are preliminary and need to be experimentally validated before being applied to display evaluations. One criterion for a good evaluation method is that the entities of the metrics should be maximally orthogonal to each other so that a compact, independent set of pertinent factors can be elucidated. The metrics presented in this report were developed under a theoretical framework: Complexity lies in the interaction between the system and the observer with the constraints of task requirements. This framework views the complexity as an entity of three orthogonal dimensions: numerical size, variety and relation. It results in a 3x3 table containing metrics of complexity: three complexity factors evaluated by the three stages of brain information processing.

In conclusion, this report presents a framework for developing metrics of information complexity in automation displays. The framework is described as follows: 1) information complexity is the combination of three basic factors: numeric size, variety and relation; 2) complexity factors are associated with the functions at three stages of brain information processing: perception, cognition, and action; and 3) the metrics of complexity can be derived by associating task requirements to brain functions. The framework incorporates many human factors studies involving interface evaluation. Within this framework, we identified a set of metrics to assess the complexity of automation displays in air traffic control. We expect that these metrics will not only be used for evaluation of new systems but will also serve as a guideline for interface design.

Acknowledgments

The author thanks Dr. Carol Manning for the valuable discussions throughout this study. I Also thank Drs. Scott Shappell, David Schroeder, Lawrence Bailey, and Earl Stein for their invaluable advices about this study. I also appreciate the effort of all the FAA internal reviewers whose comments greatly improved the quality of this paper.

References

Bennett CH (1990). How to define complexity in hhysics, and why. In: Zurek WH, ed. Complexity, entropy and the physics of information. Redwood City, California: Addison-Wesley; 137-48.

Conway AR, Engle RW (1996). Individual differences in working memory capacity: more evidence for a general capacity theory**.** Memory; 4(6):577-90.

Cowan N (2001). The Magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences; 24(1):87-114.

Crutchfield JP, Young K (1989). Inferring statistical complexity, Physics Review Letters; 63:105.

Grassberger P (1991). Information and complexity measures in dynamical systems. In Atmanspacher H, Scheingraber H, eds. Information dynamics. New York: Plenum Press; 15-33.

Halford GS, Wilson WH, Phillips W (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental and cognitive psychology. Behavioral Brain Sciences; 21(6):803-31.

Heylighen F (1989). Self-organization, emergence and the architecture of complexity. Proceedings of the 1st European Conference on System Science; Paris: AFCET; 1989:23-32.

Hopkin VD (1995). Human factors in air traffic control. Bristol, PA: Taylor & Francis.

Joseph JS, Chun M and Nakayama K (1997). Attentional requirements in a “preattentive” feature search task. Nature; 387:805-7.

Kallus KW, Barbarino M, Van Damme D (1997). Model of the cognitive aspects of air traffic control. Brüssel, Brüssel: EUROCONTROL; Ref.No. HUM.ET1.ST01.1000-Rep-02, (55 S.).

Langton C (1991). Life at the edge of chaos. In: Langton C, Taylor C, Farmer J, Rasmussen S, eds. Proceedings Artificial Life II; Redwood City, California: Addison-Wesley; 41–91.

Schmidt BK, Vogel EK, Woodman GF, Luck SJ (2002). Voluntary and automatic attentional control of visual working memory. Perception and Psychophysics; 64*:*754-63.

Scott WA (1969). The structure of natual congnition. Journal of Personality and Social Psychology; 12:261-78.

Sears AL (1994). Automated metrics for user interface design and evaluation. International Journal of Biomedical Computation; 34:149-57.

Stein ES (1992). Air traffic control visual scanning. Washington DC: Federal Aviation Administration; Report No: DOT/FAA/CT-TN92/16.

Tullis TS (1985). A computer-based tool for evaluating alphanumeric displays. Human-Computer Interaction: INTERACT; 1984 Sep; London, England.

Xing J, Heeger D (2001). Quantification of contrast-dependent center-surround interaction. Vision Research; 41:571-83.

Xing J, Manning C (2005). Complexity and automation displays of air traffic control: Literature review and analysis. Washington DC: Federal Aviation Administration; In press.

ATC OPERATIONAL ERRORS: EXCEEDING THE LIMITS OF COGNITIVE CAPACITIES

Jing Xing
Lawrence L. Bailey
FAA Civil Aerospace Medical Institute, Oklahoma City, OK 73125

Operational errors (OEs) in air traffic control are made by novice controllers and those with vast experience. Since air traffic control requires a high level of cognitive processing, this study was intended to elucidate the association of OEs and the capacity limits of cognitive processing in the brain. For this purpose we developed a memory-attention model and used the model to analyze 93 runway incursion OEs related to controller-controller communications. The analysis showed that roughly 60% of the OEs might have been associated with exceeding the capacity limits of attention and memory. We identified seven types of capacity limits as potential factors contributing to the OEs: 1) inattention; 2) attentional blink; 3) memory overload; 4) disruption of memory consolidation; 5) habit interference; 6) goal interference; and 7) similarity interference. The results suggest that controllers might be able to prevent certain types of OEs simply by being aware of those limitations. The model is preliminary and needs to be validated before being applied to a broader range of OE analysis.

Introduction

Air traffic control (ATC) provides a demanding environment in which controllers must maintain an up-to-date picture of the traffic situation. Understandably, attention and memory are critical in such tasks. Previous studies (Bales, Gilligan, & King, 1989; Cardosi, 2001) have demonstrated that controller's memory failure is one of the most significant causal factors of runway incursion operational errors (OE). Runway incursions are defined as the loss of separation between an aircraft or a vehicle on the ground with an aircraft taking off or landing. While studies established the connections between memory and OEs, it remains unclear why attention and memory would fail under the circumstances and how such failures could have been avoided.

The role of attention and memory in cognitive processing has been well established in the psychological literature, as has its capacity. Not surprising, attention and memory capacity play an equally critical role in ATC. For example, if a controller performs a task that exceeds these capacity limits, the brain may fail to process certain information, possibly leading to an OE. Thus, this report was intended to examine the relationship of OEs with the capacity of memory and attention. We developed an operational model of memory-attention processing, applied the model to the analysis of runway incursion OEs, and identified seven types of capacity limits that had the potential to cause OEs. This report presents the model and our assessment of the relationship of attention / memory capacity with runway incursion OEs.

Methods

We first developed an operational model of memory-attention processing based on extensive studies in cognition and neurophysiology. The model consists of several memory buffers; information communication between the buffers is carried out by the mechanism of attention. We then applied this model to a number of OEs. The analysis of an OE was carried out in three steps. We first mapped an OE narrative to an action description. The action description of an OE includes four parts: the initial state of the traffic situation, task requests, a sequence of actions made by controllers, and the final consequence. The second step was to map the action description to the memory-attention model. Although we did not know exactly what was in the controller's memory at the time an OE occurred, we inferred the pieces of information needed for memory and attention to carry out an action. The last step was to determine whether the memory and attention activities demanded by the actions exceeded any capacity limit of the model.

Results

An Operational Model of Memory-Attention Mechanisms

A great deal of research has been devoted to understanding memory and attention in the human brain. We generalized a model of memory-attention processing based on previous studies. The diagram of the memory-attention model is illustrated in Figure 1. The core of the model consists of two memory buffers: processing and maintenance. The operations of the model can be described as follows:

1) The processing buffer links pieces of information needed simultaneously for an action (such as landing an aircraft). It integrates information selected from

sensory inputs, long-term memory, and the maintenance buffer to make a decision. The processing buffer works dynamically. It swaps information back and forth with the maintenance buffer. Depending on the task, the processing buffer can throw away information that is no longer needed for the task or store information in the maintenance buffer for later use. By doing so the buffer empties its space for new information.

2) The maintenance buffer maintains information for a certain period of time without being attended to. The processing buffer retrieves information from the maintenance buffer when needed. Information in the maintenance buffer can be transferred to long-term memory, or simply dies if not being attended to over a period of time.

3) Attention selects pieces of information needed for an action and installs them in the processing buffer. Attention also selects the information to be transferred between the processing buffer and the maintenance buffer.

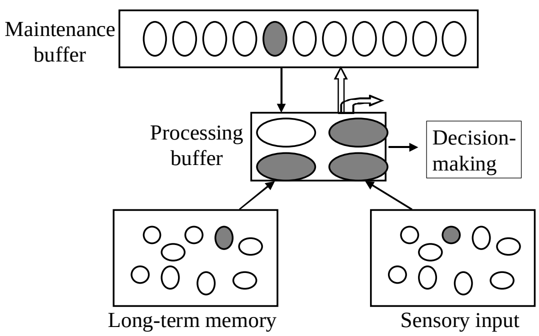


Figure 1. *The diagram of the memory-attention model.*

The most significant feature of this model that is lacking in previous studies is information swapping between the two buffers. With this feature the model can simulate dynamic cognitive activities associated with a given task. Given the task requests and the sequence of actions a controller takes, we can derive the necessary activities of the model. These activities may not be exactly the same as what actually occurs to the controller, however, they compose a minimal set of activities needed to carry out the task requests. This allows us to produce a "model performer" for the controller. By definition, a "model performer" processes information according to the memory-attention model to carry out the sequence of actions. If a model performer can perform the actions, controllers should be able to perform the action under ideal conditions. However, if the model performer cannot perform the actions, then it is expected that an ordinary controller would make errors in processing information, unless the controller develops strategies to cope with the situation. The memory buffers and attention are capacity-limited in the number of items of information they can handle and the rate at which that information can be processed. We will describe the capacity limits related to OEs in the next section.

Analysis of Operational Errors

We analyzed 93 runway incursion OEs related to local and ground controller communications that occurred between 2000-2003. The results suggested that 58 of the 93 OEs may have involved in exceeding the memory / attention capacity of the controllers. Below we described those capacity limits and provide an example OE for each type.

Memory Overload

Normal adults can only integrate about three to four items simultaneously in the processing buffer (Cowen, 2001). Beyond this limit, information in the buffer can be missed and new information cannot be processed. Moreover, overloading the processing buffer can interfere with the decision-making process. Memory overload could also occur when new information arrives for a controller at a rate that is too fast for the controller to process. Therefore, if a controller receives too many pieces of aircraft information within a short period of time, and his / her processing buffer is already occupied with other items of information, memory overload could occur. Below we present an OE that possibly involved memory overload.

Narrative: N601PA, CESSNA C172, AFTER LANDING RUNWAY 16 IS INSTRUCTED TO TURN RIGHT ON TAXIWAY M AND CONTACT GROUND CONTROL (GC). GC INSTRUCTS N601PA TO TURN LEFT ON TAXIWAY Y AND THEN LEFT ON RUNWAY 12 AND HOLD SHORT RUNWAY 16. LOCAL CONTROL (LC) TAXIES N6718Y, PIPER PA23 TO TAXI INTO POSITION AND HOLD RUNWAY 24. GC INSTRUCTS N601PA TO CROSS RUNWAY 16 AND PROCEED TO PARKING. N601PA CROSSES RUNWAY 16 AND RUNWAY 24 TO THE AMERICAN FLYER RAMP AS LC CLEARS N6718Y FOR TAKEOFF RUNWAY 24. N6718Y FLEW OVER N601PA AT THE INTERSECTION OF RUNWAY 24 AND 12.

We mapped the narrative to the action description as follows:
Initial status: Runway 12, 16, and 24 are active. Runway 12 has an intersection with Runway 24. N601PA landed on Runway 16.
Task requests: N601PA taxi to American Flyer ramp; 2) N6718Y takes off from Runway 24.
Action Sequence:
Action-1 (by LC): instructs N601PA to “turn right on Taxiway M” and “contact GC.
Action-2 (by GC): instructs N601PA to “turn left on Taxiway Y,” “left on Runway 12,” and “hold short Runway 16.”
Action-3 (by LC): instructs taxiing N6718Y to hold short Runway 24.
Action-4 (by GC): instructs N601PA to “cross Runway 16” and “proceed to parking.”
Consequences: N601PA crosses Runway 16 and Runway 24 to the ramp as LC clears N6718Y for takeoff Runway 24.

Expected activities of model performer for GC

Long-term memory: Airport / runway configurations, coordination rules.
Sensory inputs: Aircraft N6718Y, aircraft N601PA, aural communications with LC and pilots.
Maintenance buffer:Runways that are active, status of runways (in-use or not-in-use).
Processing buffer for Action-2 (by GC): N601PA, Taxiway Y, Runway 12, and Runway 16.
Processing buffer for Action-2 (by GC): N601PA, Runway 16, and parking to the ramp.

The processing buffer of the model performer for GC is overloaded with at least four pieces of information to perform Action-2. As a result, the performer cannot retrieve the information about Runway 24 to the buffer. The consequence is that the performer “forgets” to coordinate the aircraft crossing Runway 24. Therefore, we assigned memory overload as a potential causal factor to this OE.

Memory Consolidation

Attention selects pieces of information and places them in the processing buffer for decision-making. If a piece of information is to be used later, it is then transferred from the processing buffer to the maintenance buffer. The installation of new items in the maintenance buffer requires the involvement of attention, and may take several seconds to consolidate within memory. If a piece of information requires rehearsal but the attention is distracted, information can be lost. Therefore, if a controller is distracted immediately after he or she processed a piece of critical information (such as instructing a runway crossing), the information might appear to be forgotten. Below is an example OE that may have involved disruption of memory consolidation.

Narrative: ALL TIMES UTC. 1923:49 - LOCAL CONTROL (LC) TAXIED N75537 (AIRCRAFT #2) INTO POSITION AND HOLD ON RUNWAY 31 FOR TRAFFIC DEPARTING RUNWAY 26. 1924:40 - GROUND CONTROL (GC) COORDINATED WITH LC TO CROSS RUNWAY 31 AT TAXIWAY CHARLIE WITH N17323 (AIRCRAFT #1). 1924:45 - LC APPROVED N17323 TO CROSS RUNWAY 31 AT TAXIWAY CHARLIE. 1924:59 - LC CLEARED N75537 FOR TAKEOFF ON RUNWAY 31. 1925:10 - GC OBSERVED N75537 DEPARTING RUNWAY 31 WITH N17323 ON THE RUNWAY AND ADVISED LC … .

The error occurred as LC approved N17323 crossing Runway 31, and 14 sec later LC cleared N75537 for takeoff on the same runway. Once LC approved N17323 crossing, he or she may have immediately switched attention to N75537 without consolidating the memory about “Runway 31 in-use” in the maintenance buffer. As a result, the information about “Runway 31 in-use” could have slipped out of the memory system and resulted in an OE. Therefore, we assigned disruption of memory consolidation as a potential causal factor of the OE.

Inattention

Attention is required for sensory inputs to be processed in the memory system. However, the resource of attention is limited to very small spatial and temporal ranges. Hence, instead of a complete, detailed world, we only process a small part of it, the part we are attending to. In other words, you can look right at something and not see it if you are not paying attention. That is a phenomenon known as “inattentional blindness” (Simons, 2000). Similarly, there is inattentional deafness: if we don’t attend to some voices, we won’t hear them even if we listen to it. Therefore, if a controller is not paying attention to what he looks at or listens to, the sensory information could go by without the controller being aware of it. Below is an example OE that may involve inattention.

Narrative: GROUND CONTROLLER (GC) COORDINATED WITH LOCAL CONTROL TO "GO ON THE RUNWAY" WITH TWO VEHICLES AT 1212 LCL. THE VEHICLES RECEIVED CLEARANCE AND PROCEEDED ON THE

RUNWAY FROM MIDFIELD TO THE DEPARTURE END OF THE RUNWAY. THE GROUND CONTROLLER WAS RELIEVED AND INCLUDED THE TWO VEHICLES ON THE RUNWAY IN THE POSITION RELIEF BRIEFING. AT 1221 LCL LOCAL CONTROL CLEARED N3076C TO LAND ON RUNWAY 19; THE VEHICLES WERE ON THE RUNWAY. A MEMORY JOGGER (A FLIGHT PROGRESS STRIP HOLDER WITH "VEHICLE" ON IT) WAS USED PROPERLY, BUT THE LOCAL CONTROLLER DID NOT SEE IT IN HIS SCAN. THE LOCAL CONTROLLER SCANNED FINAL AND THE RUNWAY FROM THE APPROACH END TO MIDFIELD. A THIRD VEHICLE WAS AUTHORIZED BY LOCAL CONTROL TO CROSS RUNWAY 19 AT MIDFIELD.

In this narrative, the local controller landed N3076C on Runway 19 without seeing the trucks on the runway or the memory aid. The controller also authorized another vehicle crossing Runway 19 without seeing N3076C. These are typical cases of "look not see," i.e., inattentional blindness. Therefore, we assigned inattention as a potential causal factor of the OE.

Attentional Blink

Cognitive experiments demonstrated that, when a subject is told to look for two targets among many inputs presented sequentially, detection of the second of two targets is impaired if it is presented less than about one second after the first. This is known as "attentional blink" (Di Lollo, Enns, & Rensink, 2000). When a new piece of information in sensory inputs captures attention, it takes time for the attention mechanism to focus on the target and install it into the processing buffer. During that period, the attention mechanism does not respond to any new arrival of information. If the second target appears during that time, it cannot be attended to and is not reported to the memory system. Therefore, if a controller receives two or more pieces of critical information presented quickly in a sequence, attention freeze might occur, and the second or third target could get lost. Below is an OE that may involve attention blink .

Narrative: N125AJ WAS HOLDING SHORT OF RWY 27 AT B4 WAITING TO CROSS THE RUNWAY. LC INITIATED COORDINATION WITH GC TO CROSS 27 AFTER TRAFFIC ON 1 MILE FINAL. GC ACKNOWLEDGED THE COORDINATION WITH OPERATING INITIALS ONLY. GC SUBSEQUENTLY CROSSED N25AJ BEFORE TRAFFIC ON 1 MILE FINAL LANDED. GC UNDERSTOOD THE COORDINATION DONE BY LC GAVE HIM PERMISSION TO GO AHEAD AND CROSS THE RUNWAY IN FRONT OF THE LANDING TRAFFIC.

In this narrative, LC sent two pieces of target information to GC: "cross 27" and "after traffic." If the two pieces of information were spoken quickly without an interval between them, and GC did not make an effort to pay attention to the full message, then perhaps only the first target "cross 27" captured GC's attention, and it could result in attentional blink on the second target. As a result, only the information "cross 27" would enter into the processing buffer. In that case, GC would only process the information "go ahead" but not the information "after traffic."

Habit Interference

A person's knowledge and experience can interfere with the processing of sensory inputs. This is because the processing buffer receives information from both sensory inputs and long-term memory. If those two pieces of evidence are in conflict, interference occurs (Stroop, 1935). Therefore, if a controller receives a piece of information that is in conflict with what he had learned by experience, unless he / she makes an effort to suppress the response from the experience, the previous experience can lead to misinterpretation of the sensory input. Below is an example OE that appears to involve habit interference.

Narrative: WHILE WORKING GROUND CONTROL IN THE TOWER, TRUCK 24 WAS ALLOWED TO CROSS AN ACTIVE RUNWAY WITHOUT COORDINATION WITH LOCAL CONTROL. IN THIS INCIDENT, TRUCK 24 (AN AIRPORT MAINTENANCE VEHICLE) REQUESTED TO DRIVE TO A POSITION ON RWY32 (A NON-ACTIVE RUNWAY) WHICH WOULD BE ON THE WEST SIDE OF RWY19 (AN ACTIVE RUNWAY). THE ORIGINAL POSITION OF TRUCK 24 AT THE TIME OF THE REQUEST WAS ON THE EAST SIDE OF THE FIELD IN A NON-MOVEMENT AREA AND THE ONLY WAY FOR IT TO GET TO THE REQUESTED POSITION WOULD BE TO PROCEED FROM TAXIWAY C TO RWY32. THAT ROUTE CROSSES OVER RWY19. THE CONTROLLER'S INSTRUCTIONS TO TRUCK 24 WERE TO, "PROCEED AS REQUESTED AND ONCE ON RWY32, HOLD SHORT OF RWY19 AT ALL TIMES." THE CONTROLLER HAD IN HIS MIND THAT THE TRUCK WAS REQUESTING TO GO TO A POSITION ON RWY32 WHERE

MEN WERE ALREADY WORKING THAT HAPPENED TO BE ON THE EAST SIDE OF RWY19 AND WOULD NOT REQUIRE CROSSING ANY ACTIVE RUNWAYS. TRUCK 24 PROCEEDED AS INSTRUCTED AND CROSSED OVER RWY19 AT TAXIWAY C. NO PRIOR COORDINATION BETWEEN GC AND LC HAD BEEN ACCOMPLISHED.

The problem in this narrative was that the controller assumed that the truck requested to go to a position where "men were already working." The model performer of the controller could have the following information in the processing buffer: visual or auditory input about the truck's destination (west side of RWY 19) and the inferred destination (east side of RWY19) from long-term memory based on the experience that trucks usually go to working sites. In this case, the two pieces of information are in conflict. As a result of interference, the processing of the sensory input was biased by the input from long-term memory. Therefore, we assigned habit interference as a potential causal factor for this OE.

Goal Interference

People often have trouble performing two relatively simple tasks concurrently. When doing so, performance errors increase dramatically (Tombu & Jolicoeur, 2003). In the memory-attention model, the processing buffer combines pieces of information needed simultaneously for an action, thus the brain only make one action schema at a time. The brain makes different schemas for multiple task goals by switching between tasks, which swaps information back and forth between the maintenance buffer and the processing buffer. The faster the switches, the greater chance for some pieces of information to get lost or the information for different schemas to get mixed up. Therefore, if a controller tries to meet two task goals at the same time, such as landing an aircraft and coordinating a runway crossing, it is possible that the two schemas will interfere with each other and lead to an OE. Below is an example OE that may involve goal interference.

Narrative: LOCAL CONTROL EAST CONTROLLER (LCE) IS RESPONSIBLE FOR RWY 9L, 17, AND 8. THE LOCAL CONTROL WEST CONTROLLER (LCW) IS RESPONSIBLE FOR RWY 9R. THERE WERE SEVERAL DEPARTURES WAITING FOR TAKEOFF ON 9L AND 8. LCE TAXIED USA1464 INTO POSITION AND HOLD ON 9L. N6182A HAD LANDED AND CLEARED 17 AT TAXIWAY KILO. LCE SWITCHED N6182A TO THE GROUND CONTROLLER (GC) AND INSTRUCTED GC TO CROSS 9L AT ECHO WITH N6182A. LCE ALSO COORDINATED WITH LCW TO CROSS 9L WITH HIS TRAFFIC INBOUND TO THE RAMP. PRIOR TO ISSUING A TAKEOFF CLEARANCE TO USA1464, LCE ASKED IF CROSSINGS WERE COMPLETE. LCW ACKNOWLEDGED IN THE AFFIRMATIVE BUT GC DID NOT RESPOND. LCE CLEARED USA1464 FOR TAKEOFF AND AS THE AIRCRAFT BEGAN HIS TAKEOFF ROLL, LCE OBSERVED THE BE20 ENTER 9L ON TAXIWAY ECHO …

The OE occurred because LCE did not notice that GC did not respond to the inquiry about "crossing complete", and cleared USA1464 for takeoff. LCE tried to meet several task goals simultaneously and the pieces of information needed for the goals interfered with each other. The information required for the task goals included: 1) USA1464, hold on Runway 9L; 2) N6182A, cleared Runway 17, Taxiway Kilo; 3) GC, N6182A, cross 9L, Taxiway Echo, and 4) LCW, BE20, cross 9L, Taxiway Echo. The information "cleared Runway 17" and "N6182A crossed 9L on Echo" could interfere with the information about BE20 crossing 9L. As a result, LCE made the wrong conclusion that BE20 completed the crossing. Therefore, we assigned two potential causal factors to this OE, goal interference and memory overload.

Similarity Interference

Items of information can be correctly processed in one's working memory only if they can be well discriminated. Increasing similarity between the items increases difficulty of memory and the decision process (Dean, Bub, & Masson, 2001). Therefore, if a controller simultaneously, or within a short period of time, processes two or more aircraft that bear a great similarity in their call signs, interference may occur, and the controller could mistake one for the other. Below is an OE example that may involve similarity interference.

Narrative: AT 1541L, GFT9434 REPORTED ON FINAL TO RUNWAY 9R AND LOCAL NORTH CONTROLLER (LCN) CLEARED THE AIRCRAFT TO LAND. AT 1544L, LCN INSTRUCTED AES501 TO TAXI INTO POSITION AND HOLD ON RUNWAY 12. AT 1545L, LCN INSTRUCTED AAL628, USA4035 AND GFT 9434 TO CROSS RUNWAY 12. LCN HAD INTENDED TO ISSUE THE CROSSING CLEARANCE TO GFT9459, HOLDING SHORT OF RUNWAY 12, NOT GFT9434 …

One problem raised as LCN mistook "GFT 9434" for "GFT9459" when LCN instructed "AAL628, USA4035 and GFT 9434 to cross Runway 12." LCN confused the two aircraft because their call signs were similar and they were processed within the same period of time. During the incident, LCN worked with at least four aircraft for landing, crossing, and takeoff. Using the model analysis, we inferred that the processing buffer was overloaded for the model performer of LCN. Therefore, we assigned the reasons for the OE to be memory overload and similarity interference.

Statistic Results

We found that 58 out of 93 runway incursion OEs analyzed may have associated with the various types of capacity limits described above. We calculated the percentage of each type of capacity limits in the 58 OEs. Figure 2 shows the results. Along the horizontal axis are the seven types of memory-attention capacity limits. The height of the bars represents the percentage of each type in the OEs. The results show that memory overload is the most common causal factor. Notice that an OE can involve more than one type of capacity excesses. For instance, goal interference and similarity interference are often associated with memory overload. That is, when memory overload occurs, other types of capacity limit excesses are more likely to occur compared to the situations where memory load is low.

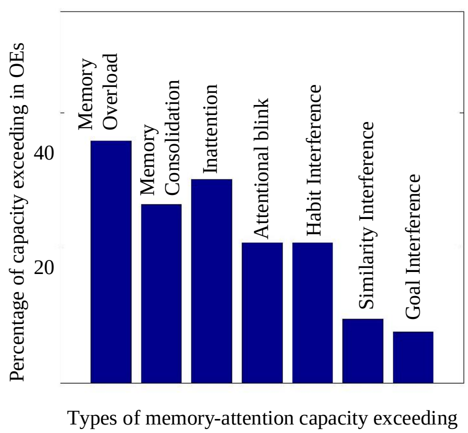


Figure 2. *The percentage of seven types of memory and attention capacity limits in the runway incursion OEs related to controller-controller communications.*

Discussion

We identified seven types of memory-attention capacity limit excesses that had the potential to lead to 58 of the 93 OEs. Among them, memory overload, disruption of memory consolidation, and inattention occurred most frequently. We need to point out that exceeding these capacity limits does not necessarily occur at a high task load. It is reasonable to assume that lack of attention more likely occur under lower task load situations, where controllers do not realize a need for intensive cognitive processing. On the other hand, even the occurrence of memory overload is not necessarily the result of a high task load. While memory overload may involve in many ATC tasks, most of time controllers could use strategies to break down the memory load. However, in situations with task load is low and moderate, controllers may not feel the pressure to handle memory overload, and thus, may tend not to use those strategies. To develop coping strategies and use them properly, one has to understand why and how capacity limits are exceeded. A previous effort made by Stein identified memory lapses occurred in ATC task (Stein, 1989). This report made a further step in attempting to elucidate the association between OEs and cognitive limits. The results need to be validated before being applied to a broader range of OE analysis.

References

Bales, R. A., Gilligan, M. R., & King, S. G. (1989). An Analysis of ATC-Related Runway Incursions with Some Potential Technological Solutions, McLean, VA: *MTR-89W21*.

Cardosi, K. (2001). Controller and pilot error in surface operations. In *Proceedings of 4th USA / Europe Air Traffic Management R&D Seminar*, Sante Fe, NM.

Cohen, N. (2001). The Magical Number 4 in Short-term Memory: A Reconsideration of Mental Storage Capacity. *Behavioral and Brain Sciences*, *24*, 1, 87-114.

Dean, M. P., Bub, D. N., & Masson, M. E. (2001). Interference from related items in object identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 733-743.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual pathways. *Journal of Experimental Psychology: General*, 129, 481-507.

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences*, 4, 147-55.

Stein, E. S. (1989). Controller memory handbook Eashington DC: The FAA technical report; No. DOT/FAA/CT-TN89/58.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 43-62.

Tombu, M. & Jolicoeur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception Performance*, 29, 3-18.

PILOTS' CONFLICT DETECTION WITH IMPERFECT CONFLICT ALERTING SYSTEM FOR THE COCKPIT DISPLAY OF TRAFFIC INFORMATION

Xidong Xu, **Christopher D. Wickens**, and **Esa M. Rantanen**
University of Illinois at Urbana-Champaign
Savoy, Illinois

Twenty-four pilots viewed dynamic encounters between the pilot's "ownship" and an intruder aircraft on a 2-D simulated Cockpit Display of Traffic Information (CDTI) and estimated the point and time of closest approach. A three-level alert system provided a correct categorical estimate of the projected miss distance (MD) on 83% of the trials. The remaining 17% of alerts incorrectly predicted MD. The data of these pilots were compared with a matched "baseline" pilots, who viewed identical trials without the aid of automated alerts. Roughly half the pilots depended on and benefited from this automation, and others did not. Those who benefited did so when problems were difficult but not when they were easy. Furthermore, automation benefits were observed only when automation was correct, but automation costs were not observed when it was in error. While assisting miss distance prediction, the automation led to an underestimate of the time remaining till the point of closest approach.

Introduction

Cockpit Display of Traffic Information (CDTI) will play a key role in a new flight environment known as free flight, allowing pilots to detect and avoid potential conflict by providing graphic information regarding nearby traffic's locations, speeds, altitudes, and other information relative to the ownship (Johnson, Battiste, & Bochow, 1999; Wickens, Helleberg, & Xu, 2002). Airborne conflict detection is a cognitively demanding task, and consequently automated aids have been invoked to assist pilots in their new charge. However, future flight paths are inherently uncertain due to a number of factors such as wind shift, pilots' intentions to change flight plans, and look-ahead time, making perfect predictions impossible (e.g., Kuchar, 2001). These uncertainty factors may lead to two types of errors in automated conflict alerting: misses (no alert of real conflict) and false alarms (safe separations treated as conflicts). In addition to the safety consequences that may result from pilots' over-trusting or over-depending on these erroneous automation outcomes, both high false alarm rates and high miss rates may cause operators to mistrust the system, which may in turn cause under-use (or under-dependence) and even disuse of the system (Parasuraman & Riley, 1997). The implications of this unreliability of automation in conflict detection are of particular interest to us.

In general, correctly functioning automation tends to improve overall system performance relative to unaided performance (Dixon, Wickens, & Chang, in press; Metzger & Parasuraman, 2005; Yeh & Wickens, 2001). However, automation benefits may not always be realized if the manually performed task had been easy (Rovira & Parasuraman, 2002), whereas automation benefits can be substantial when tasks are difficult in their manual form (Dixon & Wickens, 2004; Maltz & Shinar, 2003, Wickens & Dixon, 2005). On the other hand, costs of inaccurate automation may be larger for difficult tasks than for easy tasks (Dixon & Wickens, 2004; Maltz & Shinar, 2003; Wickens, Gempler, & Morphew, 2000). The costs and benefits can easily be interpreted with the mediating concept of automation *dependence***.** As tasks become more difficult, users become more dependent on automation to assist them, which will provide greater benefits when the automation is correct, but greater costs when it "fails" due to reduced situation awareness and/or skill degradation resulting from complacency (Parasuraman, Sheridan, & Wickens, 2000). One noticeable phenomenon is that when reliability is above 70%–75%, there are benefits but no costs relative to manual performance, especially when the task is difficult (e.g., Wickens & Dixon, 2005; Maltz & Meyer, 2003).

It appears that only two experiments have examined the issue of human responses to imperfect automation in aviation conflict detection and avoidance, by Metzger and Parasuraman (2005) and Wickens et al. (2000). The findings of Metzger and Parasuraman (conflict detection in air traffic control) and Wickens et al. (conflict avoidance using a CDTI) collectively show that correct automation is beneficial to performance and inaccurate automation poses costs, consistent with the general pattern found in other studies. The results are also in agreement with the general finding that costs and benefits are more likely to emerge for difficult (vs. easy) task, which would more likely make people depend on automation.

The goal of the present study focused on how correct and erroneous predictions of an imperfect automation alert affected performance in relation to the unaided baseline performance reported in Xu, Rantanen, and Wickens (2004), and how the effect of automation reliability was modulated by task difficulty. Based on the literature reviewed above, we formulated several hypotheses: (1) that conflict detection performance using a CDTI with an imperfect automated alerting system (83% reliable in the current study) predicting the miss distance (MD) at the closest point of approach (CPA) between the ownship and an intruder would be better than unaided performance (Wickens & Dixon, 2005); (2a) that correct automation with a valid MD alert would improve performance and (2b) error automation with invalid MD alert would hinder performance relative to manual performance on equivalent difficulty (Metzger & Parasuraman, 2005); (3a) that increasing trial difficulty would amplify the effect of reliability as mediated by increased dependence; that is, for correct automation, automation would provide greater performance *improvement* relative to manual performance for hard trials than for easy trials, and (3b) for automation errors, automation would induce greater performance *costs* relative to manual performance for hard trials than for easy trials (Dixon & Wickens, 2004).

Method

Participants

Twenty-four pilots (22 male and two female; age ranging between 18-25 years, with a mean of 19.8 years) different from the baseline study reported elsewhere (Xu, Rantanen, et al., 2004) were recruited from the Institute of Aviation, University of Illinois at Urbana-Champaign.

Simulation and Task

The CDTI depicted ownship and intruder in a map (top-down) view (see Figure 1). The display represented ownship by a white triangle and the intruder by a solid circle in cyan, yellow, or red, depending on the MD alert level. Ownship icon was positioned in the center of the display throughout the whole experiment, thus yielding an egocentric view of the traffic situation, where the ownship icon appeared to be stationary to the participant. The ownship and the intruder were flying at the same altitude on straight converging courses and at constant but not necessarily same speeds. At the start of a trial, a conflict predictor provided a three-level MD alert (no alert if MD > 3.5 nm; low level alert if 1.5 nm < MD < 3.5 nm; and high level alert if MD < 1.5 nm). The three levels of MD alert were indicated by different colors of the intruder icon, along with different verbal warnings (cyan and no verbal warning for no alert, yellow and "traffic traffic" for low level alert, and red and "conflict conflict" for high level alert). The intruder icon retained the color throughout a trial and the verbal warning was given once at the beginning of a trial. To simulate a less than perfectly reliable predictor, on one in every six trials, the automation provided erroneous prediction of MD, indicating MD that was in a greater (a miss) or smaller separation (a false alarm) category than the true value.

Participants individually observed the development of a conflict scenario for 15 sec, after which the scenario froze. They were then required to mentally extrapolate the development of the scenario, press a key when they estimated that the CPA was reached, thereby providing the estimate accuracy of time to CPA (TCPA), and move the cursor to a location that they believed was the CPA, thus providing the estimate accuracy of MD. Pilots were instructed that when the MD alert was correct, they were supposed to take advantage of it. However, when they believed that the predictor provided invalid MD prediction, the pilots were asked to ignore it and make their estimations based on their own judgments.

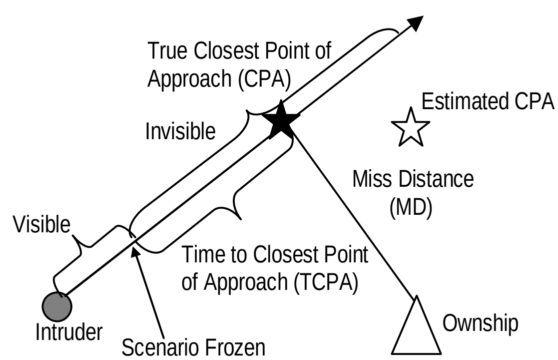


Figure 1. *Schematic illustration of key components of the experimental paradigm.*

Experimental Design and Procedure

This experiment employed a repeated measures design. However, the data for pilots in this experiment using automation were statistically compared with matched data from the "baseline" study, on identical conflict trials, performed without the aid of automation, as reported in Xu, Rantanen, et al. (2004). The trials with differing geometries in the baseline experiment that produced the easy and hard trials were randomly chosen to create an independent variable of task difficulty for the present experiment, which was varied within subjects. Trial difficulty was inferred from pilot performance, with smaller errors indicating easy trials and greater errors hard trials (see Xu, Wickens, & Rantanen, 2004 for full details).
The second independent variable was automation correctness (error vs. correct). Each participant re-

ceived 60 correct automation trials and 12 automation error trials (reliability of 83%), the latter equally representing misses and false alarms, of large and small magnitude. The 24 pilots were chosen such that their flying experience was roughly equal to that of the corresponding pilots in the baseline experiment. The 72 trials were quasi-randomly presented to the pilot. The automation error trials were in turn quasi-randomly distributed within the total 72 trials.

The participants were explicitly told that the MD predictor would not be 100% reliable. They performed ten practice trials, with a valid predictor for the first six trials and an invalid predictor for the remaining four trials, being informed explicitly of the invalidity of the last four trials. Then they participated in one experimental session to complete two blocks of 36 trials each for one to two hours in total.

Dependent Measures

The dependant variables were absolute and signed MD and TCPA estimate errors, derived by subtracting the true values from the estimated values. Absolute errors would reveal the estimation accuracy, and signed errors would reveal the estimation directions (under- or overestimate), an indication of biases. Our attention was focused on the MD measures as pilot-estimated MD represented the most safety-critical aspect of the pilot's assessment of conflict risk.

Results

Data Reduction and Analysis

A good measure of automation dependence is the difference in performance between conditions of error and correct automation (Maltz & Shinar, 2003), with a large difference being indicative of heavy dependence. We measured automation dependence by the difference in absolute MD estimate error between the automation error trials and the correct automation trials, given that only MD prediction was automated, as well as the fact that MD is the primary measure of conflict risk. The difference was calculated separately for each individual pilot, thus yielding two levels of automation dependence (light and heavy) for the 24 pilots using a median-split method. The light dependence pilots mostly encountered easy trials, whereas the heavy dependence pilots mostly had hard trials.

Analyses for Heavy Dependence Group

Hypothesis 1 was tested by two-sample t-tests for means and hypotheses 2 and 3 by 2 × 2 ANOVAs.

Overall Effect of Automation (Hypothesis 1). Absolute MD estimate error was .13 nm smaller in the current experiment (M = .33 nm) than in the corresponding trials collected in the baseline experiment (M = .46 nm), $t(22) = -1.83$, $p = .04$, suggesting that the automated alerts used here, even though imperfect, nonetheless benefited MD estimation. However, absolute TCPA estimate error did not differ significantly between the automation and manual (baseline) groups, $t(22) = .63$, $p = .27$.

MD Estimate Error (Hypotheses 2 & 3). Figure 2 presents the MD error for the baseline and automation experiments. A 2 (automation vs. manual baseline) × 2 (easy vs. hard conflict problems) mixed ANOVA for the automation *error* trials and the corresponding baseline (manual) trials (the left side of Figure 2) revealed that absolute MD estimate error did not significantly differ between the two experiments, $F(1, 22) = .56$, $p = .46$, and performance on hard trials was poorer than on easy trials in both experiments, $F(1, 22) = 17.40$, $p < .0001$. Furthermore, the same ANOVA revealed that the difference in performance between easy and hard trials in the present experiment was not reduced compared to that in the baseline experiment, since the interaction between experimental condition and task difficulty was not significant, $F(1, 22) = .31$, $p = .59$.

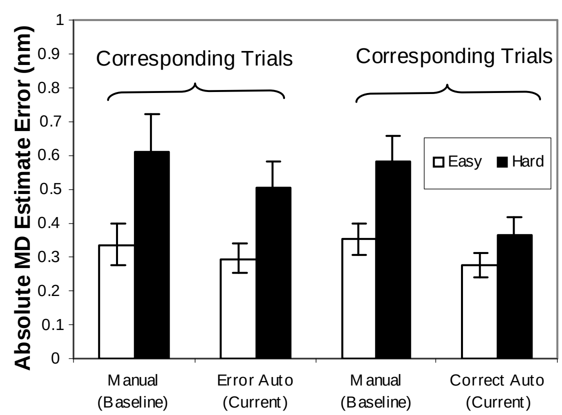


Figure 2. *Absolute MD estimate errors for heavy dependence group by automation correctness and task difficulty, and the corresponding baseline trials.*

In contrast, a 2 (automation vs. manual) × 2 (easy vs. hard) mixed ANOVA for the *correct* automation trials and the corresponding baseline trials (the right side of Figure 2) revealed that performance was better (smaller error) than in the baseline experiment, $F(1, 22) = 4.19$, $p = .053$, and performance on easy trials was better than on hard trials, $F(1, 22) = 36.73$, $p < .0001$. Most importantly, the difference in performance between the easy and hard trials in the current experiment was reduced compared to that in the baseline experiment, indicated by a significant interaction between experimental condition and task difficulty, $F(1, 22) = 6.89$, $p = .015$.

Analysis on *signed* MD estimate error suggests that there was a tendency for the MD to be less underestimated in the current experiment compared to the baseline experiment, especially when the automation was correct and the task was hard (see Xu, Wickens, et al., 2004 for detailed analysis). Therefore, the automation moved the *signed* estimates closer to the true value, reducing a conservative bias to underestimate MD that had been observed in the baseline study (see Xu, Rantanen, et al., 2004)

TCPA Estimate Error (Hypotheses 2 & 3). The results of a 2 (automation vs. manual) $\times$ 2 (easy vs. hard) mixed ANOVA revealed that when the automation was present and in *error* (left half of Figure 3), absolute TCPA (time) estimate error did not differ significantly from that in the baseline experiment, $F(1, 22) = 2.35$, $p = .14$; and performance on the hard trials was constantly poorer than on the easy ones in both experiments, $F(1, 22) = 28.86$, $p < .0001$. There was no significant interaction between the two factors $F(1, 22) = .40$, $p = .53$.

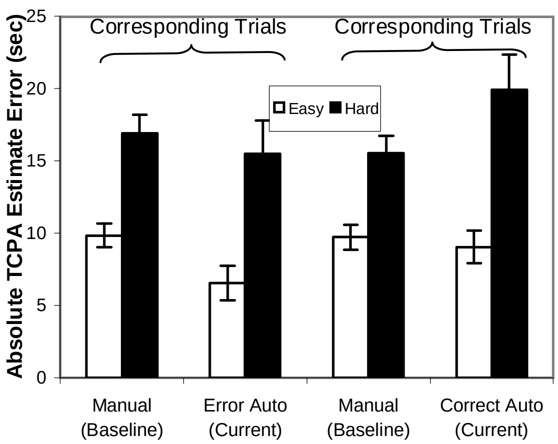


Figure 3. *Absolute TCPA estimate errors for heavy dependence group by automation correctness and task difficulty, and the corresponding baseline trials.*

However, when the automation was *correct* (right half of Figure 3), the increase in task difficulty imposed greater cost to performance than in the baseline experiment as indicated by greater absolute TCPA estimate error. This trend was confirmed by the results of a 2 (automation vs. manual) $\times$ 2 (easy vs. hard) mixed ANOVA, which revealed that the hard trials were still harder than the easy ones in the current experiment, $F(1, 22) = 69.32$, $p < .0001$, but the significant interaction between experimental condition and task difficulty suggests that the hard trials (with automation) in the current experiment induced greater TCPA (time) estimation error than the corresponding hard trials (without automation) in the baseline experiment, $F(1, 22) = 6.37$, $p = .019$.

Similar to the analyses on signed MD estimate error, we also looked at the data on *signed* TCPA estimate error (estimating conflict to be too early or too late) in both experiments. The analyses revealed that for both the correct and error automation trials, TCPA was more underestimated than their baseline counterparts, (closest passage estimated to arrive sooner than it actually would); this bias was amplified on hard trials, in both experiments (see Xu, Wickens, et al., 2004 for more detailed analysis).

Analyses for Light Dependence Group

Analyses were performed for the light dependence group in the same way as those for the heavy dependence group and the results show that none of the hypotheses were supported for the light dependence group, *F*s < 1 and *t*s < 1. This pattern of results thus suggests that the light dependence pilots did not use the automation, and hence were unaffected by its properties. Moreover, absolute MD estimate error was greater on the hard trials than on the easy trials for both the correct automation trials and the corresponding baseline trials, $F(1, 22) = 16.36$, $p = .001$, and the difference between hard and easy trials was not significantly reduced by automation, $F(1, 22) = .047$, $p = .83$, suggesting that these pilots should have used (but did not) automation for support.

Discussion

First, we had not originally anticipated the wide range of automation dependence between participants. Given such a range, it made sense to focus our hypothesis testing regarding automation properties primarily upon those who depended on automation in the first place, since those who did not would be expected to show generally null results of automation correctness (and indeed they did). Because those low dependence pilots were people who were more likely paired with those pilots who encountered easier problems in the baseline experiment, they also generally received easier problems in the current experiment. It appears that those low dependence pilots did not feel the need to obtain assistance from the automation, presumably because the task was relatively easy, although our analysis revealed that they should still have used it to improve performance. In contrast, since the high dependence pilots received the more difficult trials, it might have appeared to be an advisable strategy to depend on the automation to enhance performance. Indeed they generally were found to benefit from automation regarding the most critical safety-relevant or risk measure of conflict understanding, the estimation of MD at the closest point of approach. Performance of these pilots was better than that of their demographically matched counterparts in the baseline experiment, facing problems of equivalent difficulty but unaided, thus, supporting Hypothesis 1. Importantly, the data show that with an error rate of 17% (83% reliability), pilots clearly benefited from imperfect automation, a data point that adds to the general conclusion that imperfect automation above a 70-75% rate is better than no automation at all when workload is high and the task is difficult (Wickens & Dixon, 2005).

The analysis examining Hypothesis 2 revealed, as expected, that benefits were only realized when automation was correct and not when it was in error (thus supporting Hypothesis 2a; Figure 2). However, the results were a little surprising in that even on the automation error trials performance was no worse than its level had been in the baseline experiment, and sometimes showed a hint of being better (thus refuting Hypothesis 2b). That is, unlike other findings, erroneous automation did not yield a "complacency cost" of over-dependence, corresponding to an automation-induced beta shift (e.g., Maltz & Shinar, 2003; Metzger & Parasuraman, 2005; Yeh & Wickens, 2001). One partial explanation is that pilots were clearly pre-warned of the less-than perfect characteristics, and so were presumably not "caught" by a first failure effect, which is typically used to document the effect of over-trust, over-dependence, or "complacency" (e.g., Yeh & Wickens, 2001).

How did the high dependence pilots show a benefit from imperfect automation when it was correct, but no cost when it was wrong? Part of the answer may be because the pilots' response (positioning the cursor on the location of the projected CPA) was different from the actual guidance given by the automation predicted MD. In interpreting our results, we assume that when the high- and low-level alert appeared, pilots invested a high level of perceptual and cognitive processing of the raw data—a careful inspection—in order to most accurately estimate the CPA. This effort investment was greater than that for corresponding pilots in the baseline experiment, who did not receive the alert. Such behavior would lead to enhanced accuracy even when the alert was incorrect. When the alert was "silent" in contrast, pilots might have maintained an equivalent level of inspection to their manual baseline counterparts.

Another, parallel way of accounting for the data is to assume an overall improvement in performance of the current experiment versus the baseline experiment, perhaps due to a motivational increase from having the automation available (e.g., Beck, Dzindolet, Pierce, & Piatt, 2003). Within the overall improved performance, the cost-benefit differences associated with automation error versus correct still existed (at least on the difficult problems; see Figure 2). However, any cost for error automation was then entirely offset by the overall benefit of improved motivation and performance, particularly when the alert sounded, as described above, triggering a closer inspection of the raw data.

The finding that automation benefits emerged on high difficulty trials (thus supporting Hypothesis 3a; Figure 3) is a familiar and expected one (e.g., Dixon & Wickens, 2003, 2004; Maltz & Shinar, 2003). It is also important to note that a major feature of the high difficulty was the long distance to the closest point of approach, creating a lengthening of space over which projection must take place, that would be typical as we extrapolate the current results to the more strategic uses of the CDTI that are envisioned (e.g., 2-4 minute look-ahead time). In such a case, pilots would either have to project across a larger region of the display or if the display scale were minified, they would have to project across a slower velocity symbol movement, a prediction that is also more difficulty (Xu, Rantanen, et al., 2004) and so, again, would be likely to benefit from imperfect automation.

Another finding that was not anticipated was the distance-time estimation accuracy trade-off that was produced by automation. That is, while automation appeared to improve the accuracy of performance on the most critical task associated with conflict estimation–the estimation of miss distance at the CPA– it actually disrupted the accuracy of estimating the time till that CPA would occur. Why this occurred may be accounted for by a resource trade-off—the requirement to process both the automated alert and the raw visual data for miss distance required more resources. Such resources were diverted from the time estimation process, which was itself resource limited (Zakay, Block, & Tsal, 1999). Given then that time would be more poorly estimated as a consequence of resource diversion, pilots adopted a "conservative strategy" to underestimate that time; that is, to give themselves less time available than they really have.

In conclusion, the results have clearly illustrated the benefits that can be provided by even imperfect or "unreliable" CDTI alerting, at least given the relatively high reliability level about 80%. Such benefits –without costs – are, we believe, the result of three factors: (1) Raw data were available to be inspected; (2) pilots were calibrated to the approximate reliability level, and (3) a three-level alert was employed. We might project that increases in multi- task workload to a level more typical of the cockpit might amplify the benefits, just as decreasing the automation error rate would have had the same effects. However, it is possible that these two changes, while amplifying the benefits of correct automation, may have led to the emergence of costs on automation-error trials. Finally, caution needs to be exercised when generalizing the results here regarding the effects of automation unreliability to the real world situations, where the conflict base rate is much lower than in the present study.

Acknowledgements and Authors' Note

This research was supported in part by a grant from NASA Ames Research Center (NASA NAG 2-1535). Dr. David C. Foyle was the technical monitor. Views expressed here are those of the authors and do not necessarily reflect the views of NASA. The authors thank Ron Carbonari for the programming support. The first author is now at San Jose State University Foundation at NASA Ames Research Center, CA.

References

Beck, H. P., Dzindolet, M. T., Pierce, L. G., & Piatt, N. (2003). Looking forward: A simulation of decision aids in tomorrow's classroom. *Proc. HFES 47th Annual Meeting* (pp. 330-334). Santa Monica, CA: HFES.

Dixon, S. R., Wickens, C. D., & Chang, D. (in press). Mission control of multiple unmanned aerial vehicles: a workload analysis. *Human Factors*.

Dixon, S. R., & Wickens, C. D. (2004). *Reliability in Automated Aids for Unmanned Aerial Vehicle Flight Control: Evaluating a Model of Automation Dependence in High Workload* (AHFD-04-05/MAAD-04-1). Savoy, IL: UIUC, AHFD.

Johnson, W. W., Battiste, V., & Bochow, S. H. (1999). A cockpit display designed to enable limited flight deck separation responsibility. *Proc. 1999 World Aviation Conference*. Warrendale, PA: Society of Automotive Engineers.

Kuchar, J. K. (2001, March). *Managing uncertainty in decision-aiding and alerting system design*. Paper presented at the 6th CNS/ATM Conference, Taipei, Taiwan.

Maltz, M., & Meyer, J. (2003). Use of warnings in an attentionally demanding detection task. *Human Factors, 43*(2), 217-226.

Maltz, M., & Shinar, D. (2003). New alternative methods of analyzing human behavior in cued target acquisition. *Human Factors, 45*(2), 281-295.

Metzger, U., & Parasuraman, R. (in press). Automation in future air traffic management: Effects of reliable and imperfect conflict detection aids on controller performance and mental workload. *Human Factors*.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, and abuse. *Human Factors, 39*(2), 230-253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 30*(3), 286-297.

Rovira, E., & Parasuraman, R. (2002). *Sensor to shooter: Task development and empirical evaluation of the effects of automation unreliability*. Paper presented at the Annual Midyear Symposium of the APA, Division 10 (Military Psychology) and 21 (Engineering Psychology). Ft. Belvoir, VA.

Wickens, C. D., Gempler, K., & Morphew, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors, 2*(2), 99-126.

Wickens, C. D., Helleberg, J., & Xu, X. (2002). Pilot maneuver choice and workload in free flight. *Human Factors, 44*(2), 171-188.

Wickens, C, D, & Dixon, S. R. (2005). Is there a magic number 7: the benefits of imperfect automation. AHFD-05-01/MAAD 05-01. Savoy, IL: UIUC, AHFD.

Xu, X., Rantanen, E. M., & Wickens, C. D. (2004). *Estimation of conflict risk using cockpit displays of traffic information* (AHFD-04-11/FAA-04-4). Savoy, IL: UIUC, AHFD.

Xu, X., Wickens, C. D., & Rantanen, E. M. (2004). *Imperfect conflict alerting systems for thecockpit display of traffic information* (AHFD-04-8/NASA-04-2). Savoy, IL: UIUC, AHFD.

Yeh, M., & Wickens, C. D. (2001). Display signaling in augmented reality: The effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors, 43*(3), 355-365.

Zakay, D., Block, R. A., & Tsal, Y. (1999). Prospective duration estimation and performance. In D. Gopher, & A. Koriat (Eds.), *Attention and performance XVII—Cognitive regulation of performance: Interaction of theory and application* (pp. 557-580). Cambridge, MA: The MIT Press.