17TH INTERNATIONAL SYMPOSIUM ON AVIATION PSYCHOLOGY

MAY 6 – MAY 9, 2013

WRIGHT STATE UNIVERSITY, DAYTON, OHIO

WRIGHT STATE UNIVERSITY

For your convenience, the table of contents is linked to the body of the document and is keyword searchable.

Table of Contents

***Symposium*: High Fidelity Simulation and Aviation Training to Improve Problem Solving Skills and Coordination**

Situation Awareness

Cockpit Automation

***Symposium*: NextGen 2018: What Is and Is Not Likely to Change for Controllers**

Pilot/Crew Training

Workload Management

Tuesday Posters

Air Traffic Control Challenges

Optimizing Display Designs

Wednesday Posters

Plenary Researcher Panel

Safety

Visual/Ocular Indices in Flight

Symposium: Human-Centered Automation in Future Air Traffic Management Systems

Modeling for Aviation Safety

Surface Operations & More

Military Selection

Cognitive & Physiological Workload Indices

Symposium: Display Effectiveness & Neuroergonomics in Tactical Airlift

Controller-Pilot Communication

Pilot & ATC Selection

Auditory/Vibrotactile Displays

APPLYING SYSTEMS THINKING TO AVIATION PSYCHOLOGY

Nancy G. Leveson
Aeronautics and Astronautics Dept., MIT
Cambridge, Massachusetts

Hazard analysis is at the heart of system safety. It can be described succinctly as “investigating an accident before it happens.” A hazard is selected, such as two aircraft violating minimum separation standards or an aircraft losing sufficient lift to maintain altitude, and then the scenarios that can lead to that hazardous state are identified. Hazards are informally defined here as precursor states to accidents that the designer never wants the system to get into purposely. The resulting scenarios or potential paths to the hazard are then used to compute the probability of the hazardous state occurring or to design to either eliminate the scenarios or to control or mitigate them. Alternatively, after an accident, hazard analysis techniques can generate the potential scenarios to assist accident investigators in determining the most likely cause.

Most of the current widely used hazard analysis methods were created 50 or more years ago when the systems being built were simpler and were composed primarily of electro-mechanical components. Human operators mostly followed pre-defined procedures consisting of discrete and cognitively simple tasks such as reading a gauge or opening a valve. Failure rates and failure modes could be determined through historical usage or through extensive testing and simulation. Humans were either omitted from these calculations or were assumed to “fail” in the same way that electro-mechanical components did, that is, randomly and with an identifiable probability. Safety engineers and human factors experts existed in separate worlds: the safety engineers concentrated on the hazardous scenarios involving the physical engineered components of the system and human factors experts focused on the human operator such as training and the design of the physical interface between the human and the engineered system.

As software was introduced to increase functionality and desired system properties (such as efficiency and fuel savings), the role of the operator changed from one of direct controller to supervisor of the automation that actually flew the plane. The increasing complexity led to new types of human error (Sarter and Woods, 2008) and stretched the limits of comprehensibility for both the designers and the operators of these systems. We are now designing systems in which operator error is inevitable, but still blame most accidents on the pilots or operators. Something then is either done about the operator involved, such as fire them or retrain them, or engineers do something about operators in general such as marginalizing them further by automating more control functions or rigidifying their work by creating more rules and procedures, many of which cannot be followed if the system is to operate efficiently (Dekker, 2006).

At the same time, the hazard analysis methods were not updated sufficiently to take into account the new types of accident scenarios that were occurring and to treat the operator as an integral part of the larger system. As a result, hazard analyses often miss possible scenarios, especially those involving software or humans. To make progress, we need the psychology, human factors, engineering communities to come together to create more powerful hazard analysis methods—and therefore ways to improve the system design—that are appropriate for the systems being built and operated today. This paper describes a potential approach to doing that. It starts from an extended model of accident causality called STAMP (System-Theoretic Accident Model and Processes) that better describes the role humans and software play in accidents today (Leveson, 2012).

In the next section, STAMP and an associated new hazard analysis method called STPA are described along with the resulting implications for more sophisticated handling of humans in engineering analysis and design. Proposed changes to air traffic control (NextGen) are used as an example. Then open questions are described in which the aviation psychology community could provide important contributions.

How Are Accidents Caused?

Traditional safety engineering techniques are based on a very old model of accident causation that assumes accidents are caused by directly related chains of failure events: failure A leads to failure B which causes failure C, which leads to the loss. For example, the pitot tubes freeze, which causes the computer autopilot to stop operating (or to operate incorrectly), followed by a stall warning that is incorrectly handled by the pilots, which leads to the plane descending into the Atlantic. This chain of events is an example of an accident scenario that might be generated by a hazard analysis. The underlying model of causality implies that the way to prevent accidents is to prevent these individual failure events, for example, train pilots better in how to react to a stall warning and improve the pitot tube design.

The chain-of-events model served well for simpler systems, but our more complex, software-intensive systems are changing the nature of causality in accidents. Software does not fail randomly and, in fact, one could argue that it does not fail at all. Software is an example of pure design without any physical realization. How can an abstraction fail? It certainly can do the wrong thing at the wrong time, but almost always software related to accidents are caused by incorrect requirements, that is, the software engineers did not understand what the software was supposed to do under all conditions, such as when false readings are provided by the pitot tubes. In the same way, human contributions to accidents are also changing, with the rise in importance of system design factors, such as mode confusion, that cannot be explained totally by factors within the human but instead result from interactions between human psychology and system design. Many accidents today are not caused by individual component failure but by unsafe and unintended interactions among the system components, including the operators.

STAMP was created to deal with the new factors in accidents and to consider more than individual or multiple component failure in causal analysis (Leveson, 2012). Accidents are treated not as a chain of component failure events but as the result of inadequate enforcement of constraints on the behavior of the system components. In this case, the system includes the entire socio-technical system including company management, regulators, insurance companies, the court system, government, etc.

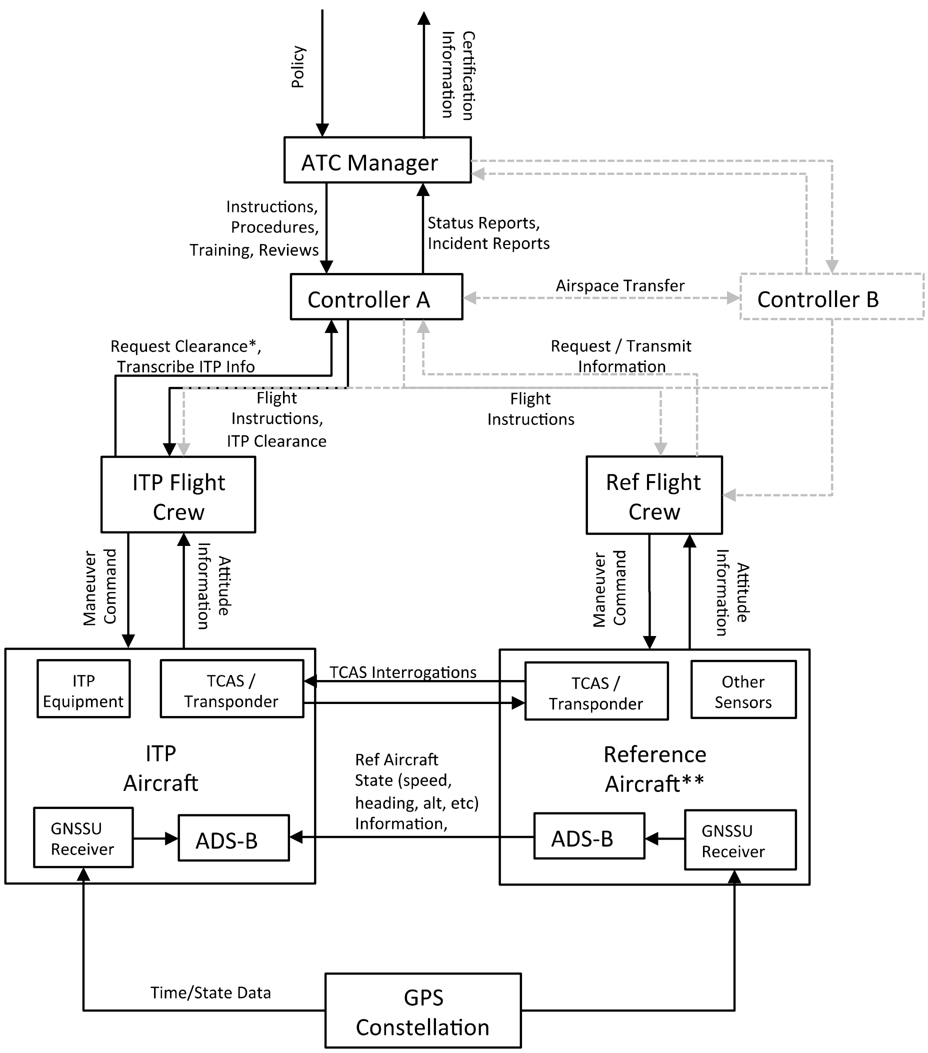


Figure 1. The Safety Control Structure for ITP

Figure 1 shows the hierarchical control structure (omitting the upper levels of management and government for simplicity) involved in a new ATC procedure called In-Trail Procedure (ITP) that allows aircraft to pass each other over the Atlantic airspace even though minimum separation requirements may be violated temporarily during the maneuver. Information about the location of both aircraft is provided through GPS and ADS-B and the ITP equipment onboard the aircraft determines whether passing will be safe at this point. If the ITP criteria for safe passing are met, the pilots can request a clearance to execute the maneuver. A hazard analysis of this system would attempt to generate the scenarios in which ITP could lead to an accident. That information can then be used by engineers and human factors experts to try to prevent accidents.

An important component of STAMP is the concept of a *process model* (see Figure 2). The safety control structure is made up of feedback control loops where the controller issues commands or control actions to the controlled process, for example, the pilot sends a command to the flight computer to ascend. In order to operate effectively, every controller must have a model of what it thinks is the state of the subsystem it is controlling. The actions or commands that the controller issues will be based at least partly on that model of the state of the system. If this model is incorrect, that is, inconsistent with the real state of the system, then the controller may do the "wrong" thing in the sense it is the right thing with respect to the information the controller has but wrong with respect to the true state of the system. If the pilots or the air traffic controller has an incorrect understanding of whether the criteria for safe execution of the ITP are met, for example, they may do the wrong thing even though they have not themselves "failed" but simply were misled about the state of the system.

The process model is kept up to date by feedback and other inputs. In humans, the process model is usually considered to be part of the *mental model*. Note that the feedback channels are crucial, both in terms of their design and operation.

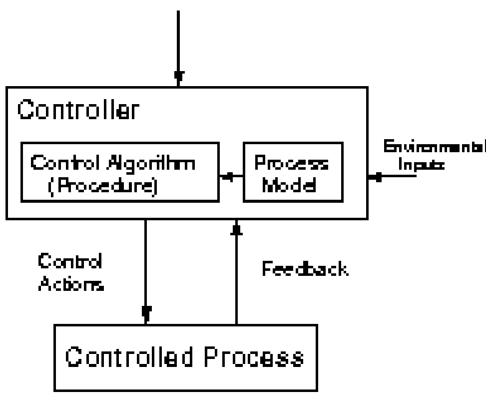


Figure 2. Every controller contains a model of the state of the controlled process

While this model works well for software and is certainly a better model for how humans work than that of random failure, it can be improved with respect to accounting for human factors in accidents. Some ideas for achieving this goal are presented later. But first, the implications of the present model are considered and the results of using it in hazard analysis are compared with traditional hazard analysis methods.

There are four types of unsafe control actions that can lead to an accident.

1. A command required for safety (to avoid a hazard) is not given. For example, two aircraft are on a collision course and neither TCAS nor an air traffic controller issues an advisory to change course.
2. Unsafe commands are given that cause a hazard. An example is an air traffic controller issuing advisories that put two aircraft on a collision course.
3. Potentially correct and safe commands are given, but at the wrong time (too early, too late, or in the wrong sequence). For example, TCAS provides a resolution advisory for the pilot to pull up too late to avoid a collision.
4. A required control command is stopped too soon or continued too long. For example, the pilot ascends as directed by a TCAS resolution advisory but does not level off at the required altitude.

Although classic control theory and control commands are emphasized here, the model is more general in terms of accounting for other types of controls on behavior than just a physical or human controller in a feedback loop. For example, component failures and unsafe interactions may be controlled through design using standard engineering techniques such as redundancy, interlocks, or fail-safe design. System behavior may also be controlled through manufacturing processes and procedures, maintenance processes, and operations. A third and important type of control over behavior comes through social controls, which may be governmental or regulatory but may also be cultural values, insurance, the legal system, or even individual self-interest. The goal of design for safety is to create a set of socio-technical safety controls that are effective in enforcing the behavior required for safety while at the same time allowing as much freedom as possible in how the non-safety goals of the system are achieved.

Identifying Hazardous Scenarios

STPA (System-Theoretic Process Analysis) is a new hazard analysis method based on the STAMP accident causation model. It works as a top-down system engineering process that starts with system hazards and then identifies behavioral constraints that must be imposed on the system components in order to ensure safety. It also assists safety analysts and system designers in identifying the set of scenarios that can lead to an accident. In practice, STPA has been found to identify a larger set of scenarios than found by traditional hazard analysis techniques, such as Fault Trees, Event Trees, and Failure Modes and Effects Analysis, particularly with respect to those scenarios involving software or human behavior.

To understand how STPA works, consider the ITA (In-Trail Procedure) example (RTCA, 2008). The STPA process first identifies the types of unsafe control actions that can lead to particular hazards and then uses that information and the control structure to identify the causes or scenarios that could lead to the unsafe control action. In the previous section, four general types of unsafe control action were identified. These are listed across the top of Table 1. The flight crew has two types of control actions they can provide (column 1): an action to execute the ITP and an action to abort it if they believe that is necessary. Within the table, the types of hazardous control actions are listed, for example, executing the ITP when the air traffic controller has not approved it or executing it when the criteria for safe passing are not satisfied. The actual process (along with automated support) to create the table are beyond the scope of this paper but the reader should be able to see easily how this could be accomplished. A complete ITP analysis can be found in Fleming, et.al. (2013).

Table 1. *Potentially Unsafe Control Actions by the Flight Crew*

Controller: Flight Crew	**Not Providing Causes Hazard**	**Providing Causes Hazard**	**Wrong Timing/Order Causes Hazard**	**Stopped Too Soon/Applied Too Long**
Execute ITP		ITP executed when not approved. ITP executed when criteria are not satisfied. ITP executed with incorrect climb rate, final altitude, etc.	ITP executed too soon before approval. ITP executed too late after reassessment.	ITP aircraft levels off above requested FL. ITP aircraft levels off below requested FL.
Abnormal Termination of ITP	FC continues with maneuver in dangerous situation	FC aborts unnecessarily. FC does not follow regional contingency procedures while aborting.		

Once the unsafe control actions have been identified, their potential causes are identified using the generic types of failures or errors that could occur in the control loop as shown in Figure 4. The information about the potential causes can then be used for system design to eliminate or reduce them, creating operational procedures, training, etc. The resulting analysis for the hazardous control action of *ITP executed when the criteria are not satisfied* is shown in Figure 5. The process model (shown in the Flight Crew box) includes such information as ownship climb/descend capability, ITP Speed/Dist criteria, relative altitude criteria, etc. A large number of potential causes for the hazardous behavior by the flight crew are identified, many of them involving reasons for the flight crew's process model to be wrong.

An important question, of course, is whether STPA is better than the traditional hazard analysis methods that are being used for NextGen. The official hazard analysis for ITP uses a combination of fault trees and event trees (RTCA. 2008). Probabilities are assigned to the various incorrect actions through a subjective process that involved workshops with controllers and pilots and eliciting how often they thought they would make certain types of mistakes. The official analysis starts with an assigned probabilistic safety objective of 1.63e-3 per ITP operation at the top of the tree. The top event, in this case, *ATC procedure is not compliant with the ITP criteria for minimum distance between aircraft*, is identified in the fault tree as being the result of potential three causes: the flight crew does not understand what the ITP minimum distance is, ATC does not receive the ITP distance but approves the maneuver anyway, or there are communication errors (partial corruption of the message during transport). Probabilities are assigned for these three causes and combined to get a probability (1.010e-4) for the top event, which is within the safety objective. The procedure is then assumed to be adequately safe, which was the goal of the analysis.

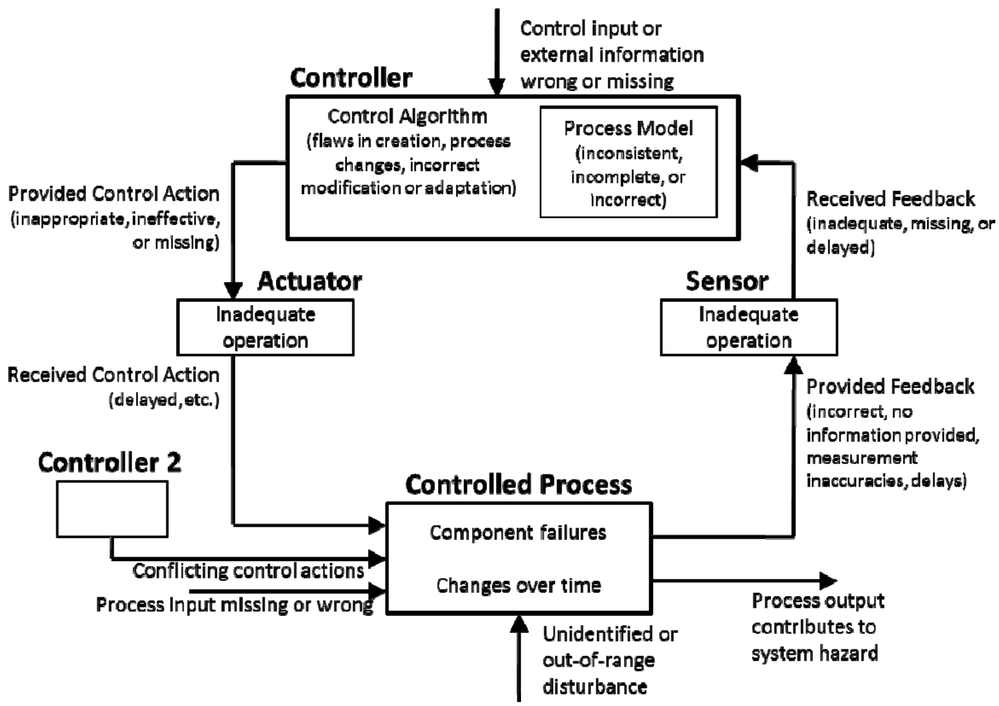


Figure 4: Generic types of problems in a general control loop that could lead to unsafe control

There are several limitations with this analysis. The probability of a human error cannot be verified, and the fault tree analysis gives no guidance on how to prevent these errors but instead assumes they happen arbitrarily or randomly. The fault tree also assumes independent behavior, however the interaction and behavior of the flight crew and ATC may be coupled, with the parties exerting influence on each other or being influenced by higher-level system constraints. Finally, the analysis asserts that communication errors are due to corruption of data during transport (essentially a hardware or software error), but there are many other reasons for potential errors in communication.

The STPA results in Figure 5 include the basic communication errors identified in the fault tree, but STPA also identifies additional reasons for communication errors as well as guidance for understanding human error within the context of the system. Communication errors may result because there is confusion about multiple sources of information (for either the flight crew or ATC), confusion about heritage or newly implemented communication protocols, or simple transcription or speaking errors. There is no way to quantify or verify the probabilities of any of these sources of error for many reasons, particularly because the errors are dependent on context and the operator environments are highly dynamic. Instead of assuming that humans will rarely “fail,” our analysis assumes they will make mistakes and specifies safety and design requirements accordingly to minimize the likelihood or the impact of human error.

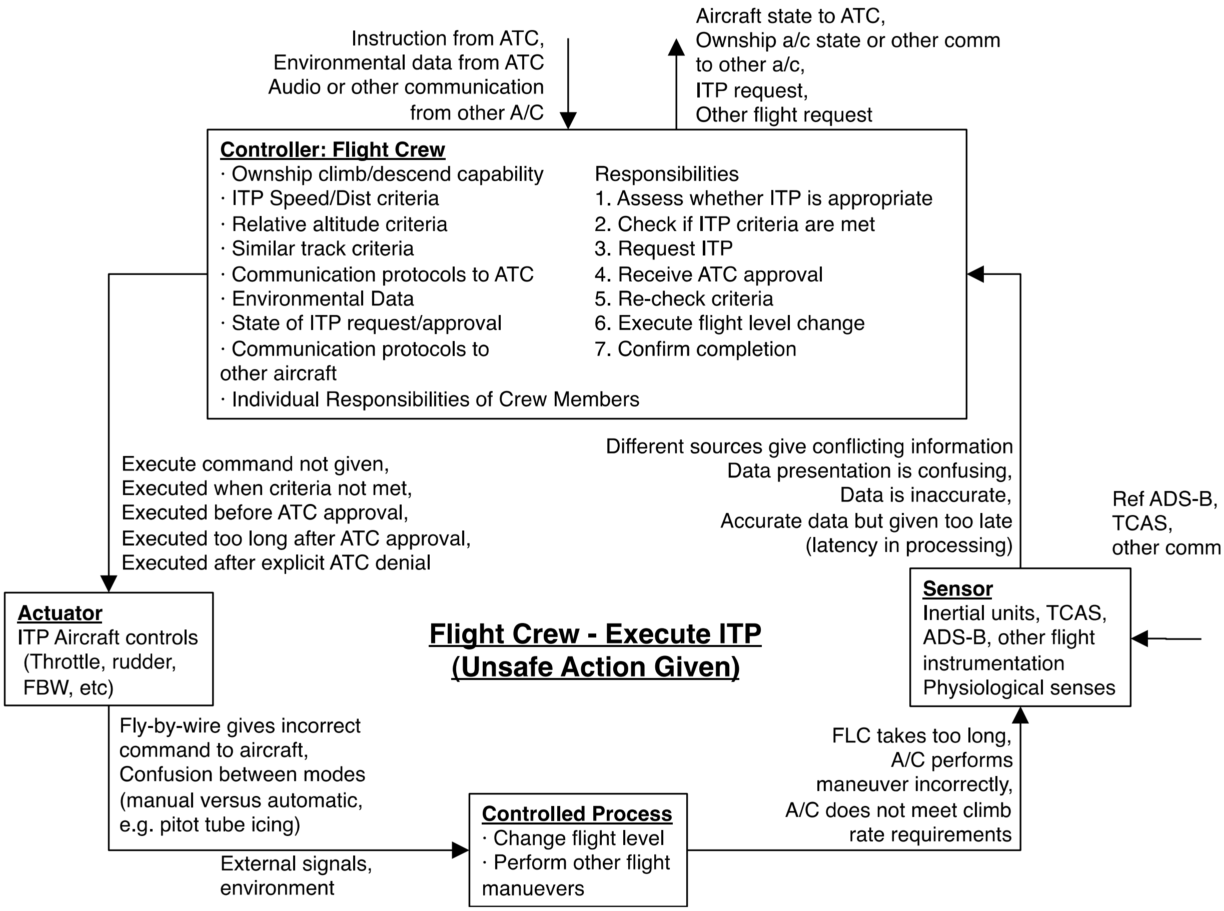


Figure 5. Conditions that could lead to the pilots executing an ITP passing maneuver when it is unsafe to do so.

Possible Extensions to STPA for Human Factors

While STPA as defined above is proving in a lot of comparative studies to be better than traditional hazard analysis techniques, it needs to be improved. The first step would be to provide a less naïve model of the human controller. While humans do not fail like mechanical components, they also do not operate with fixed algorithms (procedures) like computers as assumed above. Figure 7 shows a more realistic model of the role of humans STAMP.

There are three levels of control shown in Figure 7. The bottom two, i.e., the controlled process and an automated controller, are the same as shown previously. The top level is a first attempt at a more sophisticated model of the behavior of a human controller. Rather than having a fixed control algorithm (or procedure) that is always strictly followed, humans generate control actions using a model of the controlled process, a model of the automated controller, a model of the context in which the control is taking place as well as written or trained procedures.

Leveson (2012) has identified some basic design principles using this model to reduce human controller errors [ESW], for example, ways to support the controller in creating and maintaining an accurate mental model of the controlled process and of the automation. Known problems, such as mode confusion are included. While these principles are well known in aviation psychology, they are stated in a way that engineers can apply them directly to their designs. These principles could and should be expanded.

Another important improvement would be to extend the STPA process to include more fundamental human factors concepts. The resulting analysis could have important potential implications for providing engineers with the information necessary to design systems that greatly reduce the types of human error contributing to accidents.

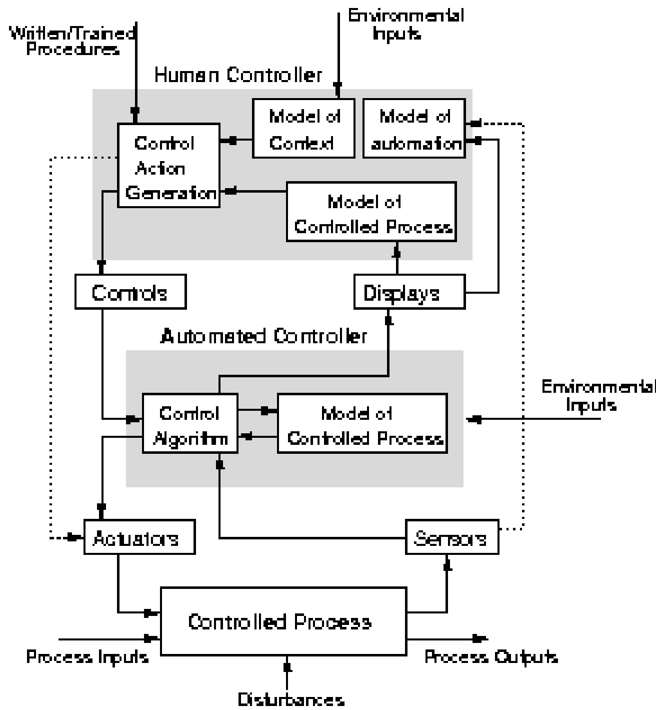


Figure 7: An Extension to Include a More Realistic Model of Human Behavior.

Conclusions

Engineering needs to get beyond greatly oversimplifying the role of humans in complex systems but the aviation psychology community needs to help them. Hazard analysis and system design techniques that were created 50 years ago are no longer useful enough. This paper has described a new, expanded model of accident causation, STAMP, based on systems thinking that could be the start for engineers and human factors experts working together to create much safer systems.

Acknowledgements

This research has been partially funded by NASA Contract NNL 10AA13C. The author also learned a great deal from conversations with John Thomas about human factors and STAMP.

References

Sidney Dekker (2006) *The Field Guide to Understanding Human Error*, London: Ashgate.

Cody Harrison Fleming, Melissa Spencer, John Thomas, Nancy Leveson, and Chris Wilkinson (2013), Safety assurance in NextGen and complex transportation systems, *Safety Science*, in press.

Nancy Leveson (2012), *Engineering a Safer World*, Cambridge: MIT Press.

Nadine Sarter and David D. Woods (1995), How in the world did I ever get in that mode? Mode error and awareness in supervisory control, *Human Factors* 37, 1, 5-19.

RTCA (2008), Safety, Performance and Interoperability Requirements Document for the In-Trail Procedure in the Oceanic Airspace (ATSA-ITP) Application, DO-312, Washington D.C., June 19.

NASA NEXTGEN FLIGHTDECK RESEARCH:
A DATABASE OF RESEARCH AREAS AND RESULTS

Kathleen L. Mosier, Alec Munc, Kendra Reich, Danielle Fox, James Swarts, Linda Tomko, Matthew O'Neill, and Kerry Cunningham
San Francisco State University

This article contains an introduction to a database created to capture important NASA or NASA-sponsored research related to NextGen flightdeck issues and operations. Documents are products of NASA's Airspace and Aviation Safety Program efforts to identify and resolve flightdeck human factors issues in NextGen, challenges to efficient operations, key areas in which technological advances are predicted to facilitate NextGen operations, research findings that can be used to develop NextGen procedures, and the potential impacts of off-nominal events.

NASA and FAA have the greatest number of research and development responsibilities for NextGen, particularly in the area of identifying and responding to human factors-related issues. Information concerning research and findings must be shared across these agencies in order to achieve a seamless integration and handoff of work to achieve the anticipated NextGen Operational Improvements. To this end, we were tasked with creating a database of NASA and/or NASA-sponsored research related to NextGen flightdeck issues and operations.

The resultant NASA NextGen Flightdeck Literature Database includes 339 documents containing the most important NASA or NASA-sponsored research related to NextGen flight deck issues and operations, produced by NASA researchers or NASA-funded researchers in the years 2006-2012. The database delineates attributes of the research in a way that can be easily searched and examined from different perspectives (e.g., by year, by topic, by researcher). It includes presentations, conference proceedings, journal articles, technical reports, and other publications, and is in the form of an Excel spreadsheet. The spreadsheet conveys relevant attributes of each document in terms of research characteristics and findings, and can be used to capture relationships between and among NASA NextGen Flightdeck Human Factors research efforts. It can also be used to identify researchers who are examining specific topics, and topics that are receiving attention in terms of NASA funding.

In the following sections, we introduce a sample of main points in topic areas such as the major changes or enhancements from current operations to planned NextGen operations, the degree of success thus far in implementing an operational concept or technology, the next steps in implementation, research findings and any identified holes in the research including missing methodologies or variables, and human factors issues and/or solutions for the NextGen flightdeck. Topic areas are catagorized as Operations, Technologies, and Human Factors Issues. Only a small number of references is included in this paper due to space limitations; however the **NASA NextGen Flightdeck Literature Database** spreadsheet can be downloaded from http://online.sfsu.edu/kmosier/

Operations

Trajectory-Based Operations (TBO)

The NASA NextGen Flightdeck Literature Database contains documents pertaining to the concept of TBO during flight, as well on surface-based TBOs. The implementation of TBOs involves a ground-based computer system with knowledge of the 4-dimensional trajectories (4D, including a time component) of nearby aircraft that would aid in scheduling and separation of those aircraft (FAA, 2009). Datalink communications technology would enable the uploading of strategic trajectories and trajectory negotiation (e.g., Coppenbarger, Mead, & Sweet, 2007; Mueller & Lozito, 2008). The articles included in this database have demonstrated the concept's potential to increase flight path efficiency and runway throughput through human-in-the-loop simulations involving both flight crews and air traffic controllers. Additionally through the manipulation of an aircraft's 4D trajectory and/or its required time of arrival at a designated waypoint, TBOs can be used to resolve potential conflicts and aid in avoidance of adverse weather conditions. One concern regarding the implementation of TBOs is the technological requirements for participating aircraft. Recent studies have demonstrated that current technologies and procedures, specifically Flight Management Systems, datalink and TCAS (Traffic Alert and Collision Avoidance System), may be sufficient for running TBOs but may not fully realize the potential for increased efficiency without additional technology. Multiple new technologies such as TBO-AID (TBO Adaptive Information Display; Bruni, Jackson,Chang, Carlin, & Tesla, 2011) and the future air navigation system (FANS) have been designed to aid flight crews in the use of TBOs without increasing workload (e.g., Coppenbarger et al., 2007); however many aircraft currently in operation are not equipped with these technologies.

Merging and Spacing (M&S)

The research on merging and spacing suggests that airport throughput can be dramatically increased over today's capacity by use of aircraft pairs (Barmore, Bone, & Penhallegon, 2009). The use of new merging and spacing techniques will also be essential in future Continuous Descent Arrival (CDA) techniques. Emphasis is being placed on pilot involvement in the process, much more so than is the case presently. Studies have found that when the flight crew is kept in the loop of merging and spacing procedures, there is a greater degree of efficiency and potential throughput for airports. Research concerning this topic area from the flightdeck perspective is relatively sparse, and different merging and spacing techniques involving the flight crew need to be explored and tested. Additionally, more controlled studies that are amenable to statistical analysis are needed, and the effect of new merging and spacing procedures on flight crew and ATC workload needs to be measured.

Departures and Arrivals

An estimated 5% increase in runway throughput can be achieved using airborne spacing techniques. Literature on arrivals and departures focuses on Area Navigation (RNAV) Departure and Arrival Procedures and CDAs (e.g., Barmore et al., 2009). Recent NASA research also focuses on integrated arrival operations along efficient descent profiles using advanced scheduling automation, tools to aid air traffic controllers, and airborne precision-spacing automation to enable fuel-efficient arrivals at busy airports during peak traffic periods. Some articles also document characteristics of aircraft that are both on and vectored from routes in the execution of area navigation (RNAV) precision departures to support precision modeling and provide for NextGen super density operations research. The Collaborative Virtual Queue (CVQ; Burgain, Feron, & Clarke, 2009) concept, which uses virtual queuing to keep aircraft away from runway queues and enable last-minute flight swapping, proposes to create departure pushback slots to enable flight departure swapping and prevent overloading the taxiway system. CVQ implementation can shorten the average departure taxiing time, reduce emissions, provide flexibility for airlines to reorder pushbacks, and increase predictability of wheels-off times by decreasing taxiway queuing. Some of the research in this area is in the operational and modeling stages. Assumed in models is that the aircraft arrive at specified locations on their prescribed paths precisely when the ATC expects them. Modeling simulations may not accurately reflect circumstances with higher traffic and poor weather conditions.

Next steps for departure and arrival studies include ongoing industry and government activities to address air-ground communication terminology, design improvements, and chart-database commonality for arrivals and departures.

Runway, Surface, and Taxi Operations

Many research efforts on runway, surface, and/or taxi operations were geared toward increasing runway throughput through improved aircraft spacing precision at landing. The Airborne Precision Spacing (Barmore, Abbot, Capron, & Baxley, 2008) concept of operations has been developed to support the precise delivery of aircraft landing successively on the same runway. Some of the latest concepts are aimed at supporting more fuel efficient and lower community noise operations while maintaining or increasing runway throughput efficiency. It has been estimated that surface surveillance information could improve optimization of departure operations by reducing emissions and the number of taxiing aircraft by 5.7%.

Taxi operations articles generally mentioned new ways to improve precision taxiing, using for instance a prototype surface automation tool (Ground Operations Situation Awareness and Efficiency Tool – GoSAFE; Verma, Kozon, Lozito, Martin, Bllinger, & Cheng, 2010) or a CVQ to shorten the average departure taxiing time. One issue in autonomous taxiing is the uncertainty about intentions of other aircraft, and some work assessed the efficacy of head-worn display (HWD) and head-up display (HUD) concepts (e.g., Arthur et al., 20008) for surface operations, to determine whether greater visibility increased situational awareness. Issues with these displays included some nausea experienced by pilots, as well as issues with latency, alignment, comfort ergonomics, color and other display rendering.

Other reseach focuses on technology and safety issues in surface operations. The Runway Safety Monitor (RSM; Jones & Prinzel, 2007), for instance, detects runway incursion conflicts and generates alerts in time for the crew to avoid collisions. Its detection algorithm has been found to be effective in reducing all types of runway incursions and eliminating the most severe incursions. The Runway Incursion Prevention System (RIPS; Jones & Prinzel, 2006)) has been designed to enhance surface situation awareness and provide cockpit alerts of potential runway conflicts in order to prevent runway incidents. However, some results indicated that most pilots were able to acquire incurring traffic looking out the cockpit windows (in VMC conditions), even before incursion alerting was activated.

Next steps in this area include implementing an alerting system for runway incursions, and refining display systems to aid the flight crews,

Closely-Spaced and Very-Closely-Spaced Parallel Runway Operations (CSPR and VCSPR)

A major issue facing NextGen operations is airport capacity. Even with the increased efficiency enabled by other proposed NextGen concepts, many airports are simply not large enough to handle the expected increase in traffic. With many of the nation's major airports located within cities expansion is not always an option. One proposed solution is to insert additional runways near or between existing ones creating (very) closely spaced parallel runways (VCSPR).

Some airports, such as San Francisco International Airport, currently conduct CSPR operations; however paired approaches are only permitted under visual meteorological conditions (VMC). Thus when weather conditions degrade, as they often do in San Francisco, the benefit of the extra runway is negated. Thus one of the major topics in VCSPR research is achieving VMC performance capabilities in instrument meteorological conditions (IMC). Multiple human-in-the-loop studies have shown that with the advancements in vision and conflict detection technology this performance goal is achievable, even with the occurrence of off-nominal events such as aircraft incursions (Verma et al., 2009). Additionally, analytical models for calculating the ultimate arrival, departure, and potential mixed operation capacity of closely-spaced parallel runways have demonstrated that the use of closely spaced parallel runways in all weather conditions can provide stable and predictable arrival capacity (Janic, 2008).

Approaches to (V)CSPRs would involve pairing aircraft with one in a slightly offset trail position; thus another major concern for (V)CSPR operations is the potential disturbance caused by the lead aircraft's wake-vortex. Multiple efforts have been made to model the behavior of the wake-vortex and establish a wake-free safe zone (Guerreiro, Neitzke, Johnson, Stough, McKissick, & Syed, 2010). HWDs and HUDs have been introduced in the context of CSPR approaches as a way to enhance situation awareness.

Off-Nominal Events

Off-nominal events pose a significant challenge to NextGen operations, and have been examined in general discussion and model development papers, human-in-the-loop simulations, Monte Carlo and other simulation techniques, meta-analyses looking at pilot performance. Methods such as Trajectory-Based Route Analysis and Control (TRAC; Callentine, 2011) have been used to model off-nominal events and recovery plans. Analyses address characteristics of the off-nominal events, situation (e.g., phase of flight), and the efficacy of new cockpit technologies such as highway-in the sky displays, datalink, or HUDs in helping pilots deal with events (e.g., Hooey, Wickens, Salud, Sebok, Hutchins, & Gore, 2009). The database includes studies looking at how pilots handle off-nominal events using enhanced and synthetic vision systems, off-nominal events in conjunction with VCSPR operations, in merging and spacing, future vehicles, and the effects of pilots' responses to off-nominal events in future trajectory based operations (e.g., Wickens, Hooey, Gore, Sebok, & Koenecke, 2009). Of particular importance are events that may occur during high-workload and high-traffic phases of flight (e.g., approach and landing, especially to VCSPRs), as they may disrupt CDAs. One caveat with some research is that pilots in many of the human-in-the-loop simulations were aware that they were going to experience off-nominal events, potentially limiting generalizability of their results to operational settings.

Technologies

ADS-B

Automatic Dependent Surveillance-Broadcast (ADS-B) is a satellite-based surveillance technology intended to enable increased capacity and efficiency by supporting enhanced visual approaches, CSPR approaches, reduced spacing on final approach, reduced separation in other flight phases, surface operations in lower visibility conditions, improved situational awareness, improved visibility, and reduced environmental impact by allowing controllers to guide aircraft into and out of crowded airspace with smaller separation standards than currently possible. Many documents in the database include ADS-B capabilities as a variable, but in a support capacity rather than as a focus of study. Those that focus specifically on ADS-B have typically looked at its impact on performance in conjunction with other display technologies such as Cockpit Display of Traffic Information (CDTI), or the Traffic Alert and Collision Avoidance System (TCAS; Romli, King, Li, & Clarke, 2008). So far, the addition of ADS-B has proven to provide small improvements over current conflict detection technologies. Aircraft performance capability was the main predictor of response time, rather than the speed or quality of the external data inputs.

ALARMS

Implementation of ALARMS (Alerting and Reasoning Management System) will consist of placing advanced sensor technologies into the cockpit to convey a large number of potentially complex alerts (Daiker & Schnell, 2010). The ALARMS technology will prioritize aircraft sensor alerts in a quick and efficient manner, essentially determining when and how to alert the pilot. The research thus far has mainly focused on the theoretical implications of the new ALARMS system and the challenges that will be associated with implementing it, as well as on creating human motor models to test different ALARMS scenarios.

Head-up Displays (HUD), Enhanced Vision Systems (EVS), Synthetic Vision Systems (SVS)

Many documents in the database address various display configurations, including but not limited to cockpit situation displays (CSD), heads-down displays (HDD), HUDs, enhanced vision systems (EVS), synthetic vision systems (SVS), head-worn or helmet-mounted displays (HMD), and external visions systems (XVS), as well as monocular and biocular displays. These new display technologies, and configurations of existing display technologies, are intended to provide increased visibility, symbology, and information for enhanced situational awareness and reduced pilot error, improvements in low-visibility operations, and overall enhanced pilot performance, particularly in terminal operations (e.g., Arthur et al., 2011). Display advancements are used to investigate Better Than Visual operations and Better Than Visual technologies for all-weather capabilities in NextGen such as below-minimum landings, suggesting potential changes in current FAA landing requirements. Experiments using these and other display technologies have been geared toward identifying pilot perceptions and characterizations of display clutter and influences of display clutter on pilot performance (e.g., Kaber, Alexander, Stelzer, Kim, & Hsiang, 2007). Some results suggest there may be a clutter "threshold" beyond which pilot performance degrades. This suggests that advanced technologies that include increasing display clutter may be counter-productive, which points to the need for both eliminating clutter and improving the salience of critical symbology and information. Several articles were overviews of new technologies rather than experiments or investigations into their utility and effectiveness, suggesting a need for further investigation of new flight deck display technologies, including issues such as readability in daylight (color, brightness, contrast) and disorientation and illusion issues.

Conflict Detection and Resolution (CD/R)

A major research focus for CD/R is improving the algorithms of conflict detection tools in order to create more effective vertical and horizontal resolutions with fewer secondary conflicts. Pilots will become increasingly more active in the conflict resolution decision-making process; however this may be at the cost of increasing their workload. As the amount of air traffic increases, creating conflict resolutions that avoid secondary or cascading conflicts is becoming more of a concern, and this is reflected in the amount of research that is dedicated to this concept. Some research focuses on pilot acceptance of CD/R automation (e.g., Battiste et al., 2008). New algorithms have been generally successful at creating more effective and efficient conflict resolutions with more accurate predictions of future conflicts and recovery from loss of separation (e.g, Butler & Munoz, 2009). The next step will be to test the new algorithms in more diverse and dynamic environments. A consistent drawback of most of these studies is that certain variables (aircraft location, altitude etc.) were held constant in order to test the experimental variable. Therefore, the algorithms may not perform as well in a more dynamic and realistic scenario.

Haptic Control

The main feature of haptic control technology is the ability of a control surface or system to provide tactile feedback to the pilot. The additional feedback has been demonstrated to improve pilot situational awareness of aircraft state and overall pilot performance (Goodrich, Schutte & Williams, 2011). Additionally, pilots seem to prefer the Haptic Flight Control System to traditional systems. The research regarding the haptic control has been promising with human-in-the-loop simulations showing positive effects.

Human Factors Issues

Attention

Much of the research addressing attention for NextGen applications concerns noticing and perceiving events in a situation. The N-SEEV (Noticing - Salience, Effort, Expectancy, Value) model has been a primary means to examine human attention on the flight deck, and has demonstrated success in predicting variance in pilot response to off-nominal situations (e.g., McCarley et al., 2007). Another area of attention research concerns checklist monitoring and checking, exploring factors that cause pilots skip or miss items (Dismukes & Berman, 2010). Additionally, the relationship between attention and pilot engagement and fatigue has been investigated through brain imaging techniques that provide feedback on how much attention the pilot is paying, as well as the level of fatigue experienced.

Roles and Responsibilities

Implementation of NextGen will entail collaborative decision making between air and ground, and some reallocation of roles and responsibilities. Flight crews are expected to have increased responsibility for flight paths, especially with respect to spacing and separation from other aircraft. With the expected increase in air traffic, the potential increase in ATC workload figures to be a limiting factor in the number of aircraft the system can handle. One proposed solution to this problem is to assign some of the spacing responsibilities to the flight crews themselves. Assigning aircraft to self-separation is seen to be an effective solution to the inevitable increase in traffic, particularly when there is some flexibility (Idris & Shen, 2010), lowering ATC workload substantially while maintaining an acceptable workload for the flight crews (Johnson et al., 2010). Human-in-the-loop and computer simulations have demonstrated that self-separation is able to accommodate 2-5x increases in traffic in en route operations. Self-spacing has also demonstrated the capability to increase runway throughput and facilitate the use of continuous descent arrivals through increasing spacing precision and arrival accuracy. Increased efficiency due to self-separation also is projected to decrease noise and emissions. A major obstacle to the self-separation concept is the accuracy (or lack thereof) of wind forecasts. The lack of realistic wind forecast errors was a significant limitation in the existing research. Additionally, implementation of self-separation will entail additional training, enhanced crew resource management, and tailored procedures within the flight deck.

Operator Performance

Operator performance is a large area that covers all aspects of pilot behavior and encompasses human factors variables such as workload, situation awareness (SA), and decision-making. A common methodological trend in this area is the modeling of pilot performance in these areas, particularly with respect to new technologies or off-nominal events. NextGen operations such as CDA and technologies such as EVS/SFS or other displays typically focus on increasing SA and maintaining manageable workload (e.g., Johnson et al., 2010). Other research in this area includes human-in-the-loop simulations as well as meta-analyses that compare the results of the model against previously published articles. In the future the topic of operator performance will become increasingly important as we try to predict pilot behavior under new NextGen operational conditions.

Acknowledgements

This work was funded through NASA Cooperative Agreement NNX10AK52A to San Francisco State University. Barbara Burian was the technical monitor for the literature database project.

References

Arthur J. J. III., Prinzel, L. J. III., Bailey, R. E., Shelton, K. J., Williams, S. P., Kramer, L. J., & Norman, R. M. (2008). *Head-worn display concepts for surface operations for commercial aircraft.* (NASA/TP-2008-215321). Langley, VA: NASA.

Arthur, J. J., Prinzel, L. J., Williams, S. P., Bailey, R. E., Shelton, K. J., & Norman, R. M. (2011). Enhanced/synthetic vision and head-worn display technologies for terminal maneuvering area NextGen operations. Proceedings of the *SPIE Defense, Security, and Sensing Conference* (pp. 25-29). Orlando, FL.

Barmore, B., Abbott, T., Capron, W., Baxley, B. (2008, September). *Simulation Results for Airborne Precision Spacing Along Continuous Descent Arrivals*. Presented at the 26th Congress of International Council of the Aeronautical Sciences, Anchorage, AK.

Barmore, B., Bone, R., & Penhallegon, W. (2009). Flight deck-based merging and spacing operations. *Air Traffic Control Quarterly, Vol. 17(1) 5-37.*

Battiste, V., Johnson, W. W., Dao, A. Q., Brandt, S., Johnson, N., & Granada, S. (2008). Assessment of flight crew acceptance of automated resolution suggestions and manual resolution tools. *Proceedings of the 26th International Congress of the Aeronautical Sciences*, Anchorage, AK: ICAS.

Bruni, S., Chang, A., Carlin, A., Swanson, L., & Pratt, S. (2012). Designing an adaptive flight deck display for trajectory-based operations in NextGen. In S. J. Landry (Eds.), *Advances in human aspects of aviation* (23-32). Boca Raton, FL: CRC Press.

Burgain, P., Feron, E., & Clarke, J. (2009). Collaborative virtual queue: Benefit analysis of a collaborative decision making concept applied to congested airport departure operations. *Air Traffic Control Association Institute, Incorporated, 17(2)*, 195-222.

Butler, R. W., & Munoz, C. A. (2009). *Formally verified practical algorithms for recovery from loss of separation* (NASA TM-2009-215726). Hampton, VA: NASA Langley Research Center.

Callantine, T. (2011, August). *Modeling Off-nominal Recovery in NextGen Terminal-area Operations.* Presented at the 2011 AIAA Modeling and Simulation Technologies Conference, Portland, OR

Coppenbarger, R. A., Mead, R. W., & Sweet, D. N. (2007). Field evaluation of the tailored arrivals concept for datalink-enabled continuous descent approach. *Proceedings of the 7th AIAA integrations and Operations Conference (ATI),* Belfast, Northern Ireland: AIAA.

Daiker, R., & Schnell, T. (2010). Development of a human motor model for the evaluation of an integrated alerting and notification flight deck system. *Proceedings of the MODSIM World 2009 Conference and Expo* (pp. 157-162), Hampton, VA.

Dismukes, R.K. & Berman, B. (2010). *Checklists and monitoring in the cockpit: Why crucial defenses sometimes fail.* (NASA Technical Memorandum-2010-216396). Moffett Field, CA: NASA Ames Research Center.

FAA (2009). *NextGen Mid-Term Concept of Operations for the National Airspace System, Version 1.0*. Washington, DC: FAA Air Traffic Organization NextGen & Operations Planning, Research & Technology Development, Air Traffic Systems Concept Development.

Goodrich, K. H. Schutte, P. C., & Williams, R. A. (2011). Haptic-multimodal flight control system update. *Proceedings of the 11th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, Virginia Beach, VA.

Guerreiro, N., Neitzke, K., Johnso, S., Stough, H., McKissick, B., & Syed, H. (2010). Characterizing a wake-free safe zone for the simplified aircraft-based paired approach concept. *Proceedings of the AIAA Atmospheric and Space Environments Conference*, Toronto, Ontario, Canada: AIAA.

Hooey, B. L., Wickens, C. D., Salud, E., Sebok, A., Hutchins, S., & Gore, B. F. (2009). Predicting the unpredictable: Estimating human performance parameters for off-nominal events. *Proceedings of the 15th International Symposium on Aviation Psychology,* Dayton, OH: Wright State University.

Idris, H., & Shen, N. (2010). Improving separation assurance stability through trajectory flexibility preservation. *Proceedings of the 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, Fort Worth, ID: AIAA.

Janic, M. (2008). Modelling the capacity of closely spaced parallel runways using innovative approach procedures. *Transportation Research. Part C: Emerging Technologies, Volume 16, Issue 6.*

Johnson, W., Ho, N., Battiste, V., Kim-Phuong, V., Lachter, J., Ligda, S., ... & Martin, P. (2010). Management of Continuous Descent Approach During Interval Management Operation. *Digital Avionics Systems Conference.*

Jones, D. R., & Prinzel, L. J. (2006, October). Runway incursion prevention for general aviation operations. In 25th Digital Avionics Systems Conference, 2006 IEEE/AIAA (pp. 1-12). IEEE.

Jones, D. B., & Prinzel, J. III. (2007). Cockpit technology for prevention of general aviation runway incursions. *Proceedings of the 14th International Symposium on Aviation Psychology,* Dayton, OH; Wright State University.

Kaber, D. B., Alexander, A., Stelzer, E., Kim, S-H., & Hsiang, S. (2007). *Psychophysical Modeling of Perceived Clutter in Advanced Head-Up Displays* Technical presentation at the 2007 NASA Aviation Safety Technical Conference. St. Louis, MO.

McCarley, J., Wickens, C., Sebok, A., Steelman-Allen, K., Bzostek, J., & Koenecke, C. (2009). Control of Attention: Modeling the Effects of Stimulus Characteristics, Task Demands, and Individual Differences. NASA Final Report, ROA 2007, NRA NNX07AV97A (2007).

Mueller, E., & Lozito, S. (2008). *Flight Deck Procedural Guidelines for Datalink Trajectory Negotiation.* Paper presented at the American Institute of Aeronautics and Astronautics (AIAA) Aviation Technology, Integration, and Operations (ATIO) Conference, Anchorage, AK .

Romli, F. I., King, J. D., Li, L., & Clarke, J. P. (2008). Impact of Automatic Dependent Surveillance – Broadcast (ADS-B) on Traffic Alert and Collision Avoidance System (TCAS) performance. *Proceedings of the AIAA Guidance, Navigation and Control Conference and Exhibit,* Honolulu, HI: AIAA.

Verma, S., Kozon, T., Lozito, S., Martin, L., Ballinger, D., & Cheng., V. (2010). Human factors of precision taxiing under two levels of automation. *Air Traffic Control Quarterly*, *18*(2), 113-141.

Verma, S., Lozito, S., Ballinger, D., Kozon, T., Panda, R., Carpenter, D.... Resnick, H. (2009). Evaluation of triple closely spaced parallel runway procedures for off-nominal cases. *Proceedings of the Eighth USA Europe Air Traffic Management Research and Developement Seminar,* Napa, CA.

Wickens, C. D., Hooey, B. L., Gore, B. F., Sebok, A., & Koenecke, C. (2009). Identifying black swans in nextgen: Predicting human performance in off-nominal conditions. *Human Factors, 51(5)*, 638-651

SEQUENTIAL REQUIRED TIME OF ARRIVAL INTERVALS UTILIZING EXISTING TECHNOLOGIES

Amy L. Alexander and Thomas L. Teller
MIT Lincoln Laboratory
Lexington, MA

Four-dimensional trajectory-based operations (4D TBO) requiring aircraft to meet specified timing constraints at designated waypoints along their route are a key strategy for increasing airspace throughput and efficiency. The Required Time of Arrival (RTA) function of the aircraft Flight Management System (FMS) and the ground-based Traffic Management Advisor (TMA) were tested for supporting such operations. Two different methods for implementing appropriate time intervals between sequential RTA aircraft were explored in this paper: a fixed 90-second interval versus discrete intervals varied to match the prevailing conditions. Both methods yielded a good balance between separation assurance and flow efficiency. Three different types of FMS were included in this study; the results validated the performance of the FMSs to achieve the assigned RTA clearance time, although there were some issues with meeting the altitude and airspeed restrictions specified at the meter fix, likely due to the need to actively manage speed on descent given current FMS functionality. Human factors issues in need of further examination are presented.

A key strategy for increasing National Airspace System throughput and efficiency, as presented in the NextGen Avionics Roadmap (JPDO, 2011), relies on four-dimensional trajectory-based operations (4D TBO). 4D trajectories require aircraft to meet specified timing constraints at designated waypoints along their route. Transitioning from strictly distance-based operations to include time-based standards is expected to reduce controller workload (by decreasing the use of vectors, speed commands, and holding), improve the predictability of arrival times, increase throughput, reduce time and distance flown, and reduce fuel burn (e.g., Klooster, Del Amo, & Manzi, 2009). The purpose of this work is to demonstrate the potential benefits of sequential 4D trajectories utilizing existing technology, both in the air and on the ground. The aircraft Flight Management System (FMS) is an airborne tool capable of computing a 4D trajectory and providing time guidance control to any point in the flight plan via the Required Time of Arrival (RTA) function. The Traffic Management Advisor (TMA) is a ground-based tool that can schedule arrival times at metering fixes. De Smedt and Berz (2007) examined single RTA flights under various operational conditions utilizing existing FMS functionality and pointed to the need to assess RTA applications in the context of an arrival sequence in which both time and lateral spacing will be critical to TBO.

Prior work (Teller, 2011) has demonstrated that it is challenging to implement appropriate time intervals between sequential RTA flights (or between RTA and non-RTA aircraft) that preserve adequate lateral separation while avoiding “gaps” in the arrival flow because the relationship between time interval and the resulting spacing varies with operating conditions. Assignment of sequential RTA flights at one-minute intervals, particularly in conditions yielding lower groundspeeds (e.g., headwinds), required heightened controller monitoring, more frequent interventions to assure separation, imposed higher workload for pilots and controllers, and increased the risk of RTA cancellation. On the other hand, assignment of sequential RTAs at intervals of two minutes or more yielded adequate separation between sequential aircraft at the expense of some loss of flow efficiency, particularly in conditions yielding higher groundspeeds (e.g., tailwinds). The first issue explored in this work was the use of 90-second intervals between sequential RTA flights to determine if this interval provides a good balance between separation assurance and flow efficiency without having negative impacts on workload. As an alternative to the assignment of RTAs at fixed time intervals (i.e., 90 seconds), varied discrete interval assignment based on prevailing conditions was also explored to assess whether this method might yield an adaptive balance between separation assurance and flow efficiency. This study was further designed to validate the performance of three different FMSs to achieve assigned RTA clearance times and meet applicable altitude/speed restrictions.

Method

The simulation was conducted at the MITRE/CAASD Integrated Demonstration and Experimentation for Aeronautics (IDEA) Lab. The simulated environment was driven by the NASA Multi-Aircraft Control Simulation (MACS) software application configured to emulate Denver Air Route Traffic Control Center (ZDV), specifically sectors 9, 16, and 15 (consistent with typical operations at ZDV, sectors 16 and 15 were combined). MACS was interfaced to NASA Research-TMA (R-TMA) emulation to provide terminal scheduling to the meter fix and runway in support of the research objectives. Six RTA flights were flown by three expert pilots operating FMS emulations (two FMSs per pilot) across 11 runs covering a variety of operating conditions in a within-subjects design. Two runs were conducted using fixed 90-second intervals between RTA aircraft, one in prevailing headwind conditions and one in prevailing tailwind conditions. Nine runs were conducted using discrete intervals that varied to match the prevailing conditions, namely headwinds versus tailwinds and metering versus non-metering conditions. Intervals were varied in one run by use of a controller look-up aid indexed to groundspeed and in all other runs by direct application of TMA which determines Scheduled Times of Arrival (STAs) for sequential aircraft with appropriate spacing for separation and efficiency. Variations in operating conditions (i.e., metering and traffic level/complexity) and FMS parameters (i.e., Cost Index) were examined but are not discussed below as they did not impact the results. Procedures and communication protocols were based on those developed through prior RTA work.

Equipment

Three different types of aircraft FMSs were emulated in this simulation: (1) GE Aviation sFMS used in the Boeing B737NG, (2) Aerosim/Honeywell A320 FMST, and (3) Aerosim/Honeywell B757 FMST (see Figure 1). Four instances of the B737 FMS were included; this FMS was designed to execute RTAs to any point in the flight plan route and time with a programmable time-error tolerance of 6-30 seconds. One instance of the A320 FMS was tested; this FMS was designed to actively recalculate speeds to achieve RTAs while in cruise prior to top of descent with no operator control of time-error tolerance. Two instances of the B757 FMS were used; this FMS was not designed to adjust speed during climb or descent to achieve the RTA.

Figure 1. RTA page of GE Aviation FMS in B737 (Left), Aerosim/Honeywell FMS in A320 (Center), and Aerosim/Honeywell FMS in B757 (Right).

Two controller participants used standard DSR 2Kx2K displays, computer keyboards, and mouse or trackball. The systems were configured to emulate two ZDV sectors (i.e., sector 9 and sector 16/15) and were linked by an internal lab communications network to emulate the assigned frequencies for each sector. Traffic Management Unit (TMU) was equipped with Plan Graphical User Interface (P-GUI) and Timeline Graphical User Interface (T-GUI) displays, keyboard, and mouse configured to emulate TMA functionality. Two pseudo pilots operated background traffic using a multi-feature simulation pilot interface integrated into MACS.

Scenarios

Each scenario was of 45-60 minutes duration, varying with configuration; the active flight time for each RTA flight was 35-45 minutes. All scenarios involved arrivals via the northeast corridor to Denver International Airport (KDEN). As shown in Figure 2, there are three sectors serving northeast arrivals: Sector 09 is a high-altitude sector that adjoins Minneapolis ARTCC (ZMP), includes the TMA freeze horizon, and contains the initial segment of the SAYGE6 Standard Terminal Arrival Route (STAR); Sector 16 is a high altitude-sector that incorporates the middle portion of the STAR and sets up sequencing; Sector 15 is a low-altitude sector that includes the SAYGE meter fix and the handoff to Denver TRACON. Consistent with typical operations at ZDV, sectors 16 and 15 were combined. The arrival flow was merged via SAYGE6 with restrictions in altitude (FL190; tolerance: ±100 ft) and speed (250 knots; tolerance: ±10 knots) at SAYGE.

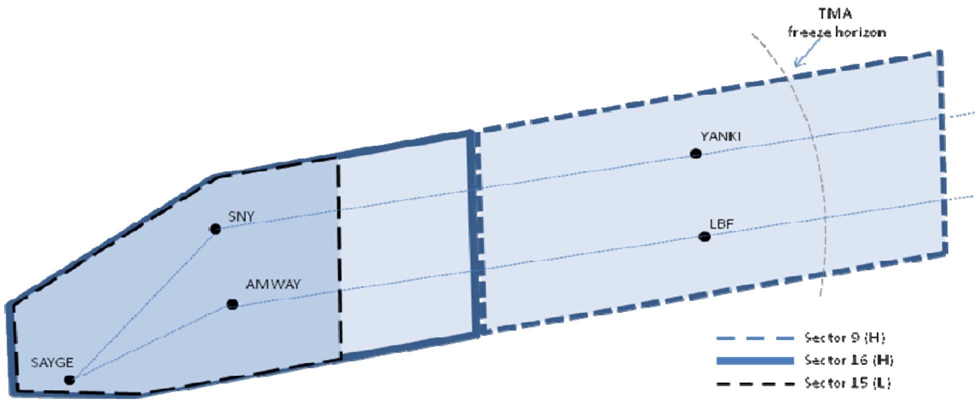


Figure 2. ZDV northeast arrival corridor sectors and SAYGE meter fix. Adapted from: SAAB Sensis Corporation 3D PAM Project.

The simulated airspace was populated by background traffic extracted from Enhanced Traffic Management System (ETMS) historical data to yield traffic flows typical to ZDV. Simulation runs were launched via pre-programmed scenarios which initialized all radar controller (R-side) positions with ZDV airspace/maps and quickly populated traffic sample flights distributed through the airspace. Additional traffic sample flights were pre-programmed to “flow in” to maintain appropriate traffic levels throughout the scenario. Environmental wind samples were extracted from historical National Oceanic and Atmospheric Administration (NOAA) Rapid Update Cycle (RUC) wind data to represent typical local wind patterns. This wind sample was used for the “headwind” runs and flipped 180° for the “tailwind” runs. For some runs, FMS forecast winds were intentionally “mis-matched” by 10 knots to validate the robustness of the FMS RTA functionality under realistic field conditions.

TMU was employed for all scenarios to calculate schedules for the prevailing environmental and traffic conditions that would deliver aircraft to specified meter points and the runway at a rate determined to ensure smooth operation, while preserving appropriate spacing. The TMA adaptation was setup in accordance with the prevailing practice at ZDV, except that the system was set to a single-runway configuration due to the limited scope of the traffic sample and data runs that investigated only the northeast arrival corridor. The Traffic Management Coordinator (TMC) adjusted constraints in TMA to generate delays according to scenario criteria and adjusted the acceptance rate, TMA matrix, and/or stream class settings as needed.

Results

One run with six flights assigned RTAs at fixed 90-second intervals was conducted in headwind conditions. Table 1 presents the assigned RTA for each flight, the actual time at which each flight crossed SAYGE, and the time difference between the RTA and actual crossing time (i.e., whether each flight was on time, early, or late). Only one flight, F4 (the A320 FMS), did not meet its RTA within the generally-accepted tolerance of ±30 seconds. While all flights met the speed restriction at SAYGE within ±10 knots, the final two flights crossed at too high of an altitude (F4: 20,214ft and F5: 19,300ft). Planned versus actual time intervals for each of two sequential flights are presented, along with the resultant planned versus actual spacing. Lateral spacing between sequential flights was in excess of the 5NM minimum spacing prescribed for the en route environment (Planned: Mean 8.8NM, Range 8.5-9.2NM;

Actual: Mean 9.0NM, Range 5.4-11.4NM). Even with the interval compression due to the early arrival of F4, the actual lateral spacing still met minimum separation standards.

Table 1.
Crossing Times, Intervals, and Spacing at SAYGE Under Headwind Conditions.

ID	Type	RTA (mm:ss)	xSAYGE (mm:ss)	Δt* (mm:ss)		Interval (mm:ss)			Spacing (NM)		
						Planned	Actual	Δi*	Planned	Actual	Δs*
F1	B737	18:33:30	18:33:12	-00:18							
					>	01:30	01:42	00:12	8.7	9.9	1.2
F2	B737	18:35:00	18:34:54	-00:06							
					>	01:30	01:37	00:07	9.2	9.9	0.7
F6	B757	18:36:30	18:36:31	00:01							
					>	01:30	01:29	-00:01	8.5	8.4	-0.1
F3	B737	18:38:00	18:38:00	00:00							
					>	01:30	00:55	-00:35	8.9	5.4	-3.5
F4	A320	18:39:30	18:38:55	-00:35							
					>	01:30	02:02	00:32	8.5	11.5	3.0
F5	B757	18:41:00	18:40:57	-00:03							

*Negative values of the time difference (Δt) indicate that the flight crossed early. Positive values of the time difference (Δt) indicate that the flight crossed late. Negative values of the interval and spacing differences (Δi and Δs, respectively) indicate that the interval was smaller than planned. Positive values of the interval and spacing difference (Δi and Δs, respectively) indicate that the interval was larger than planned.

One run with six flights assigned RTAs at fixed 90-second intervals was conducted in tailwind conditions, although one flight (F3) was terminated at the start of the run due to a software error. Table 2 presents the assigned RTA for each flight, the actual time at which each flight crossed SAYGE, and the time difference between the RTA and actual crossing time (i.e., whether each flight was on time, early, or late). All flights met their assigned RTAs within the ±30 second tolerance, all flights met the prescribed altitude restriction within ±50 ft, and only one flight exceeded the speed restriction by ±10 knots (F2: 286 knots). Planned versus actual time intervals for each of two sequential flights are presented, along with the resultant planned versus actual spacing, except for sequential intervals associated with F3. Spacing was greater in this run compared to the previous run given the increased groundspeeds associated with tailwind conditions (Planned: Mean 9.6NM, Range 9.0-10.2NM; Actual: Mean 8.9NM, Range 7.2-10.0NM). All spacing intervals well exceeded minimum separation standards.

Table 2.
Crossing Times, Intervals, and Spacing at SAYGE Under Tailwind Conditions.

ID	Type	RTA (mm:ss)	xSAYGE (mm:ss)	Δt* (mm:ss)		Interval (mm:ss)			Spacing (NM)		
						Planned	Actual	Δi*	Planned	Actual	Δs*
F1	B737	19:59:54	20:00:04	00:10							
					>	01:30	01:23	-00:07	10.2	9.4	-0.8
F2	B737	20:01:24	20:01:27	00:03							
F4	A320	20:06:18	20:06:37	00:19							
					>	01:30	01:12	-00:18	9.0	7.2	-1.8
F5	B757	20:07:48	20:07:49	00:01							
					>	01:30	01:33	00:03	9.6	10.0	0.4
F6	B757	20:09:18	20:09:22	00:04							

* Positive values of the time difference (Δt) indicate that the flight crossed late. Negative values of the interval and spacing differences (Δi and Δs, respectively) indicate that the interval was smaller than planned. Positive values of the interval and spacing difference (Δi and Δs, respectively) indicate that the interval was larger than planned.

Five runs of six flights each were conducted with discrete intervals under headwind conditions. Four runs of six flights each were conducted with discrete intervals under tailwind conditions. Table 3 presents a summary of the planned versus actual intervals and spacing for each of these nine runs.

Table 3.
Summary of Discrete Sequential RTA Intervals/Spacing at SAYGE.

Run	Conditions	Planned Intervals (mm:ss)	Actual Intervals (mm:ss)		Planned Spacing (NM)	Actual Spacing (NM)	
		Mean	Mean	Mean Δi	Mean	Mean	Mean Δs
R1	Headwinds	01:12	01:22	00:32	7.2	8.1	3.2
R2	Tailwinds	01:36	01:35	00:05	9.6	9.4	0.5
R3	Headwinds	01:40	01:39	00:09	9.6	9.6	0.9
R4	Tailwinds	01:29	01:28	00:09	8.4	8.3	0.8
R5	Headwinds	01:31	01:37	00:12	8.5	9.1	1.1
R6	Headwinds	01:20	01:13	00:26	7.4	6.9	2.5
R7	Tailwinds	01:11	01:06	00:12	6.9	6.3	1.2
R8	Tailwinds	01:23	01:24	00:11	8.5	8.7	1.1
R9	Headwinds	01:18	01:19	00:11	7.2	7.2	1.0

A 2 (sequential RTA) x 2 (winds) x 3 (FMS) multivariate analysis of variance (ANOVA) was conducted on the crossing time difference (Δt) data. Results revealed no significant differences in the method of the sequential RTA (fixed vs. discrete) or wind direction (headwinds vs. tailwinds). The type of FMS, however, was found to have a significant effect on crossing time performance. The A320 FMS (Mean 28.6s) exhibited greater error in achieving its RTA than the B737 (Mean 9.1s) or B757 (Mean 7.6) FMSs ($F(2,3)=13.7, p<0.05$). It is important to note, however, that the mean crossing time difference for the A320 FMS still fell within the accepted ±30 second tolerance. Two 2 (sequential RTA) x 2 (winds) repeated measures ANOVAs were conducted on the planned versus actual interval (Δi) and planned versus actual spacing (Δs) data (FMS was not included as a between-subjects measure as the order of flights varied across runs). 9.3% of flights did not meet the altitude restriction of ±100 ft at SAYGE while 7.4% of flights did not meet the speed restriction of ±10 knots. There were no significant differences between RTA method or wind direction.

Discussion

RTA operations provide the possibility to sequence traffic at a much earlier state than currently possible. Synchronization of the optimal trajectory, within air traffic control and airspace constraints, provides predictability not currently available, promoting time-coordinated operations and providing the potential to enable more flight-efficient operations. Two different methods for implementing appropriate time intervals between sequential RTA aircraft were explored in this paper: a fixed 90-second interval and discrete intervals varied to match the prevailing conditions. Three different types of FMS were also examined; the results validated the performance of the FMSs to achieve the assigned RTA clearance time, although there were some issues with meeting the altitude and airspeed restrictions specified at the meter fix, likely due to the need to actively manage speed on descent given current FMS functionality. The implementation of the alternatives for sequential RTA flights studied had no impact on FMS performance.

The data demonstrated that fixed 90-second intervals did yield a good balance between separation assurance and flow efficiency and were relatively straightforward to schedule (in TMA) and assign (by the controller). The use of a 90-second interval between RTAs would seem to represent the best option for a *fixed* *interval* between sequential RTAs and is a workable balance between the deficiencies of the one-minute interval (inadequate lateral spacing in conditions yielding low groundspeeds) and the two-minute interval (excess spacing in high groundspeeds, resulting in loss of flow efficiency). It should be noted that time errors, even those within tolerances, i.e., ±30 seconds, can exacerbate these spacing issues, but even under adverse cases (e.g., lead aircraft late and trail aircraft early) the 90-second interval still provides a good margin for separation assurance with comparatively low need for controller intervention. The controller participants concurred that it is likely that controllers could/would use fixed 90-second intervals between sequential RTAs in practice, provided that the TMA adaptation and meter list were configured for half-minute intervals.

As an alternative to assignment of RTAs at fixed time intervals, assignment at discrete intervals varied to match the prevailing conditions was explored to assess whether this alternative might yield an adaptive balance between separation assurance and flow efficiency. The data demonstrated that assignment of RTAs at discrete intervals did yield a good balance of separation assurance and flow efficiency adaptive to the prevailing conditions and intrinsically avoided inadequate lateral spacing (in conditions yielding low groundspeeds) and excess spacing and loss of flow efficiency (in conditions yielding high groundspeeds). The controller participants concurred that it seems a natural and simple extension of TMA to use these discrete times for RTAs, and it is very likely that controllers could/would use such a tool in practice to calculate and assign discrete times/intervals, provided that the TMA adaptation and meter list were configured for discrete time intervals and suitable symbology were provided to discriminate RTA from non-RTA aircraft.

This experiment identified potential human factors issues associated with RTA operations in five areas: tools and methods for conformance monitoring by pilots and controllers; tools and methods for assessing and predicting separation for time-based clearances; standardization of RTA-linked FMS functionality; workload impacts for pilots managing RTAs in the context of additional cockpit duties (e.g., checklists, approach planning and briefings, coordination with cabin crew and airline operations); and tools to enhance situation awareness for pilots involved in closely-spaced arrivals. Tools such as Traffic Alert and Collision Avoidance System (TCAS) and/or Cockpit Display of Traffic Information (CDTI) designed to aid pilot situation awareness, conformance monitoring, and separation assurance for RTAs are being explored in other work.

Acknowledgements

This work was sponsored by the Federal Aviation Administration under Air Force Contract #FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government. The authors would like to thank the FAA Advanced Concepts & Technology Development Office, Technology Development & Prototyping Division (ANG-C5) for sponsoring this work and Matt Modderno for serving as the FAA Project Manager. We would also like to thank MITRE/CAASD for hosting this experiment and providing FMS resources, particularly Paul MacWilliams, Niamh Lowry, Robin Kirkman, Tom Schafer, Roland Sgorcea, Will Symionow, Carmen Villani, and Mitch Wynnyk for their technical and pilot/air traffic expertise. Finally, we would like to thank Lou Rosgen and Steve Torzone of Veracity Engineering for serving as air traffic and pilot subject matter experts, and Alan Bell of Systems Enginuity for serving as a pilot subject matter expert and providing safety management systems expertise and statistical data analysis support.

References

De Smedt, D. & Berz, G. (2007). Study of the required time of arrival function of current FMS in an ATM context. Proceedings of the 26th Digital Avionics Systems Conference (pp. 1.D.5-1 – 1.D.5-10). IEEE.

Joint Planning and Development Office (JPDO, 2011). *NextGen avionics roadmap* (Version 2.0). Retrieved from http://www.jpdo.gov.

Klooster, J. K., Del Amo, A., & Manzi, P. (2009). Controlled time-of-arrival flight trials. Presented at the *Eighth USA/Europe Air Traffic Management Research and Development Seminar*. ATM.

Teller, T. L. (2011). *4D FMS TBO pilot-controller human-in-the-loop simulation* (ATC Project Memorandum No. 42PM-TCAS-0086). Lexington, MA: MIT Lincoln Laboratory.

REVIEW OF PILOT PERFORMANCE AND PILOT-AUTOMATION INTERACTION MODELS IN SUPPORT OF NEXTGEN

Angelia Sebok & Christopher D. Wickens
Alion Science and Technology
Boulder, Colorado

Computational models of aircraft pilot performance will gain importance over the next decades, as major evolutions in the national airspace continue to emerge with the NextGen program. Evaluating new technology, or procedures such as self-separation, requires time and resource-consuming pilot-in-the-loop (PITL) simulations. Models can augment PITL findings and they can help to constrain the scope of PITL simulations. If they are validated, such computational models may actually answer some design questions in place of PITL simulations. This paper summarizes a review of modeling efforts to address pilot performance, and elaborates on pilot-automation interaction models.

In the transition to NextGen operations, a major concern is identifying and evaluating potential concepts well before they are put into operation. One approach is through the use of computational modeling to predict operator performance, or human performance modeling. Computational modeling provides a means of predicting performance and evaluating numerous "what if" situations, and is thus particularly useful for evaluating to-be-built systems. In addition, models need to make valid predictions of operator performance. This requires comparing model predictions with actual human performance in real or (for to-be-built systems) simulated conditions.

Methods

To identify the scope of existing pilot performance models and their associated validation efforts, we searched more than 40 potential sources (e.g., the Human Factors Society Proceedings, Human Factors Journal, International Symposium for Aviation Psychology Proceedings, International Journal for Aviation Psychology) to identify papers that described either a modeling effort for predicting pilot performance, or an empirical study to validate such a model. Initially, we identified approximately 500 papers. Upon closer inspection, we were able to eliminate approximately two thirds of these papers as duplicates of other articles, air traffic control (ATC; not pilot) related, or model descriptions without provision of specific model predictions. This left a final set of 187 references. We reviewed these papers to characterize the modeling efforts and compare across the studies. We identified a set of criteria by which to evaluate the models, including descriptive and evaluative features. Descriptive features include the name and type of model (e.g., simulation, analytical), the specific aspect of pilot performance that was modeled (e.g., pilot-automation interaction, communication, error). Evaluative features included 1) whether or not empirical, PITL data were provided to validate the model predictions, 2) whether the empirical data provided correlations (or other quantitative evaluations), or were qualitative in nature, 3) the participants in the study (e.g., professional pilots, college students), 4) the test bed (e.g., flight simulator, desktop flight simulator with mouse and keyboard, or other). The descriptive features allowed us to distinguish the different modeling efforts, and the evaluative criteria provided data to compare the extent to which validation studies had been performed.

After this initial assessment, we conducted five separate deep dive analyses to examine in detail the state of the art of modeling efforts of particular relevance to NextGen operations: Pilot-automation interaction (PAI), error, workload and multitasking (see a companion paper, Wickens & Sebok, 2013, for details), situation awareness, and roles and responsibilities. The deep dive analyses identified how the models predict pilot performance and included a review of the verification and validation efforts for these specific topic areas. As implicitly stated previously, validation was considered to be a comparison of model predications against data gathered in an empirical PITL study. In contrast, verification efforts included subject matter expert (SME) reviews of model predictions for "sensibility" of results, or researchers' own interpretation of the results.

Overview of Model Review Results

We carried out an extensive analysis of the extent to which each modeling effort ***in the deep dive analysis*** had been verified and, if verified, also validated; by assigning ratings to the levels of verification and validation (See

Wickens et al., 2013 for details). From this analysis, we concluded that only 5% of modeling efforts included no discernible attempt at verification or validation. Nearly 40% of efforts included verification efforts, and 31% included qualitative evaluations of predictions against empirical data. Twenty-five percent of modeling efforts included human in the loop simulation studies to provide empirical data for comparison with model predictions. This situation leaves room for improvement, and the analysis identifies clear gaps. Studies frequently focused on a small subset of model predictions, rather than the full range of predictions.

Pilot-Automation Interaction Models

Overview

In NextGen operations, new technologies and capabilities are required to provide a significantly increased volume of operations. One of the key features envisioned to enable integration of these capabilities into the aviation system is a greater reliance on automation. As pilots' tasks expand to include maintaining separation from surrounding aircraft, negotiating trajectories with ATC, and monitoring weather and wake vortex conditions, automation is expected to provide pilots with the support needed to perform these tasks. Thus, models that predict pilot performance when using different types of flight deck automation are highly relevant to NextGen operations. Since "pilot-automation interaction on the flight deck" is a broad area, modeling efforts evaluated specific aspects of the domain. The models reviewed here were diverse, some focusing on specific equipment such as the Flight Management System (FMS), some on particular high-workload phases of flight, and others on specific action sequences used in programming automation.

Architectures Used for Pilot-Automation Interaction

We distinguish a modeling architecture as a software tool and / or a theoretical framework that serve as the basis for specific modeling efforts. A successful validation of one modeling effort provides support for the underlying architecture, but it does not "validate" that architecture. If an architecture addresses a specific aspect of performance (e.g., visual scanning), and one model using the architecture has been shown, through comparison with empirical data to be valid, this finding does lend support for other, similar models developed using that same architecture. Three architectures appeared repeatedly in the review of PAI models, described below.

The Adaptive Control of Thought – Rational (ACT-R) is a unified theory of cognition that integrates theories of attention, cognition, and motor actions (Anderson & Lebiere, 1998). ACT-R models cognition through "production rules" or goal-directed behavior, implemented through a series of "if-then" rules. It includes perceptual inputs and motor outputs. ACT-R's main components are modules, buffers, and a pattern matcher. ACT-R uses perceptual-motor modules (visual and manual modules) to simulate interaction with the physical environment. ACT-R models simulate declarative and procedural knowledge. In addition ACT-R models the actual time required for cognitive steps (e.g., retrieving an item from declarative memory) or implementing an action (e.g., shifting gaze, selecting an item on a display). Thus it readily models procedural activities such as programming an FMS.

The CASCaS (Cognitive Architecture for Safety Critical Task Simulation) architecture, like ACT-R, is a cognitive architecture, which provides a structure and set of rules for simulating human cognition (Lüdtke et al., 2009). CASCaS divides cognitive processes and errors into three different levels, depending on operator experience with a particular task: the autonomous, the associative, and the cognitive level which correspond, respectively, to the skill-based, rule-based, and knowledge-based levels of behavior defined by Rasmussen (1983).

The SEEV and N-SEEV models (Wickens & McCarley, 2008; Steelman-Allen et al., 2011) of visual scanning and noticing predict that attention within a given visual field (e.g., a flight deck) is driven by bottom-up factors of display ***salience*** and ***effort*** (distance between displays), and the top-down factors of ***expectancy*** (bandwidth) and ***value*** or importance of the display for the task. SEEV predicts visual scanning behavior. N-SEEV (Noticing SEEV) uses SEEV to predict scanning and **noticing** of discrete events within the cockpit.

Modeling architectures and efforts that address PAI take a variety of approaches in predicting pilot performance. Models predict performance based on pilot visual scanning and noticing, the time required to complete tasks, workload, and automation induced errors. PAI models have also been applied to design tools and proposed as a basis for adaptive automation systems. These predictions and application areas are described below.

Modeling Approaches to Predicting Performance

Attention / Noticing and Visual Scanning. Several modeling efforts addressed PAI in terms of predicted noticing of important changes on the flight deck (e.g., flight mode annunciator indications on the FMS), or in terms of visual scanning. These efforts all start with the premise that a pilot has to notice an indication to be able to interpret and respond to it, so noticing is a necessary (but not sufficient) first step that a pilot must perform. Boehm-Davis et al. (2002) used an ACT-R model to predict pilot noticing of automation mode changes. The model predicted that pilots were more likely to notice mode changes that were initiated by the pilot, rather than the automation. The authors note that similar trends were observed in previously gathered empirical data.

The SEEV and N-SEEV models of visual attention and noticing have been used to predict pilot noticing on the flight deck (Sebok et al., 2012). The SEEV and N-SEEV models have been empirically validated in previous efforts (Wickens et al., 2008; Steelman-Allen et al., 2011). In each of these efforts, model predictions in aviation flight deck contexts, including a high fidelity Boeing 747-400 simulator (Sarter et al., 2007) were found to predict empirical data of scanning and noticing behavior with correlations above 0.60.

CASCaS was used to predict visual scanning behavior, dwell times in areas of interest on an advanced FMS, and the time required to notice specific visual indications in the cruise and approach phases of flight (Lüdtke et al., 2012). The authors indicated that the *overall* correlation between model predictions and empirical data was high (r^2=0.85). For specific aspects of pilot performance, the model predictions were reasonably accurate. Predicted average dwell times on a display in three phases of flight closely paralleled by empirical data. Similarly, average noticing times for specific visual indications in two phases of flight were predicted to be approximately 1 s in each phase, and were found to be 0.8 s and 1.2 s. These results indicate that the model does a reasonable job of approximating pilot behavior. One concern is that validation data provided are for highly specific tasks or visual areas, yet the operational context includes many tasks and areas.

Polson and Javaux (2001) present a model that predicts why pilots do not often scan the flight mode annunciators, a major issue in FMS monitoring. They apply a Goals, Operators, Methods and Selection rules (GOMS; Card et al., 1983) modeling analysis that, among other features, highlights differences in task priorities in multi-tasking, to predict why this task should be of lower priority when other sources of redundant, equivalent information are available. The authors describe a qualitative evaluation of the similarity between their predictions and the data on FMS monitoring by Huttig, Anders & Tautz (1999).

Time to Complete Tasks. CASCaS was also used to predict time to to handle an uplink from ATC in the cruise and approach phases of flight (Lüdtke et al., 2012). The model predicted that the uplink would require approximately 1 minute, with slightly longer times in the approach phase than in cruise. An empirical study of those conditions revealed that pilots performed the uplink faster during approach than in cruise. No quantitative data were provided. Discussion with pilot SMEs provided insights into the reversal between model predictions and data, explaining that, during the approach phase, pilots typically have to work faster just to get everything done.

Manton and Hughes, 1990, developed a regression equation, based on previously-gathered empirical data, to predict the time to complete tasks using a Multi-Function Keyset (MFK) on the S-70B-2 Seahawk Helicopter, used by the Royal Australian Navy. The MFK, much like an FMS, includes a special purpose keyboard and an 8-line alphanumeric display, used to enter data into or view data contained in a tactical database. The equation predicts time as a function of the number of key presses required, operator pauses, and page changes. Using a stepwise regression, the authors found that the equation predicts 79% of the variance in the data ($p < 0.001$). The authors propose that the model can be used to evaluate different types of automation and system configurations.

Air Man-Machine Integration Design and Analysis System (MIDAS, v1; Pisanich & Corker, 1995) was used to predict which type of FMS automation pilots would use to perform a descent based on the time available to implement the clearance and the modality in which the clearance was delivered (voice or datalink). Three types of automation were considered: an autoload capability (the most highly automated), a CDU, and an MCP (mode control panel, the least automated). The Air MIDAS model predicted that the less time available to implement a clearance, the more likely pilots were to use a less-automated mode. Further, the model predicted that pilots were more likely to select the less-automated modes if a clearance was given by voice than by datalink. While the model was validated against a PITL simulation, the results of that validation could not be easily interpreted because of the use of inappropriate t-test statistics.

Workload. Another approach to predicting pilot performance with automation uses workload. Gil et al., 2009, used enhanced (E)-GOMS to model pilot performance when working with a flight control panel (FCP), a control display unit (CDU) or an enhanced CDU. They predicted workload based on the complexity of the procedures, including the number of submethods being performed, the number of steps needed to complete the submethods, the chunks of information that pilots needed to remember, and the number of information transactions. As complexity increases, so does workload. The authors ran the model for each of the three types of automation and collected data on the complexity indices that varied across automation types. In an empirical study, they gathered four different measures of workload: heart rate, subjective workload (NASA-TLX predictions), vertical flight path deviations, and lateral flight path deviations. They calculated the Spearman correlations for the different complexity indices and empirical performance data, and identified positive and significant ($p \leq 0.05$) correlations between the model predictions and heart rate, and between model predictions and vertical flight path deviations.

Automation-Induced Errors. CogTool (John et al., 2009) is based on ACT-R code, and models the time to complete tasks, errors made on task steps, and failure to complete task steps. In their research, the authors identified three sequential tasks associated with entering a landing speed into the CDU, a critical interface between the pilot and the FMS. They ran their model to predict errors, and iterative improved the model. CogTool accesses a latent semantic analysis (LSA) corpus of terms to predict if pilots will understand the terminology on the CDU. During their first model run, they identified that no pilots would be able to complete the first step of the procedure because they did not understand the terms. The LSA corpus represented a college student's knowledge, not the specialized knowledge that a pilot would possess. By switching to an aviation-specific corpus, the researchers obtained a 10 percent success rate on the first task. A series of other changes were implemented to account for pilots' specialized knowledge, and the model eventually predicted success rates of 92% for the entire procedure. This was considered reasonably accurate, based on one of the author's experience as a pilot who trains new pilots to use the FMS, but it was not validated against PITL simulation data.

Schoppek & Boehm-Davis (2004) used ACT-R to create a model (ACT-Fly) to model pilot awareness, cognition, and errors. They evaluated pilot use of automation at the end of the cruise phase of flight until the initial approach fix. ACT-Fly predicted when pilots would choose a more automated mode (VNAV) or a less-automated mode (FLCH and V/S) in two scenarios. In summary, the model predictions were not well supported by empirical findings. Two scenarios showed 20% and 60% agreement in terms of predicted mode selection. The model did predict the types of errors pilots would make (errors of omission and commission), but the model incorrectly predicted error recovery. Actual pilots were able to recover from their errors. Gil et al. (2009) indicate that their E-GOMS model can predict error, by identifying when the number of chunks to be held in working memory exceeds 5. This approach is based on limitations of working memory (Miller, 1956). No error predictions were made, however.

CASCaS was used by Lüdtke et al., (2009) to predict cognitive errors such as Learned Carelessness, which occurs when pilots routinely perform procedures with multiple steps included to ensure safety criteria are met. If these steps typically do not identify safety concerns, pilots learn that they improve efficiency by skipping these steps. The problem is that sometimes these unsafe conditions do exist, and, by skipping those steps, pilots may sacrifice safety for efficiency. Lüdtke & Osterloh (2010) used CASCaS to predict learned carelessness in a flight re-planning procedure. The researchers modeled a flight condition in which a pilot was repeatedly given ATC clearances that required verification. The model predicted that the pilot would, over time, begin neglecting these checks. An actual pilot performed the same conditions, and – as the model predicted – quit performing the verifications. However, unlike the model, the pilot resumed checking after receiving a supposedly related prompt. The researchers updated their model to include contextual factors (strengthening or inhibiting associations between elements in memory). The updated model then correctly predicted (in 23 of the 24 trials) when the pilot checked the vertical view.

Applications of PAI Models within Design Tools. Three papers described efforts to use PAI models in computerized design tools. These efforts used different types of models, but all had the same goal of helping aviation designers identify and avoid potential design problems. One effort (Gonzales-Calleros et al., 2010) evaluated the FMS interface design for adherence to human factors standards such as font type and color contrast between text and background. The paper outlined an approach to include a cognitive model of pilot performance, but the model was not actually integrated with the evaluation tool.

The Automation Design Advisor Tool (ADAT; Sebok et al., 2012) evaluates and compares potential FMS designs. This effort included multiple analytic models to assess design quality based on human factors principles. The analytic models evaluated design issues of 1) information layout, 2) noticeability of changes, 3) meaningfulness of terms, 4) confusability of terms and symbols, 5) complexity of system design (e.g., modes), and 6) procedures necessary to program the FMS. ADAT included attention models (Wickens & McCarley, 2008), described above, to predict pilot scanning behavior and noticing of FMS mode changes.

A third design tool, CogTool (John et al., 2009), allows a designer to create a "use-case storyboard" with a graphical user interface (GUI), and predict time to complete task or errors made. The GUI is connected with an underlying cognitive model, so the planned sequence of actions on the interface is associated with steps such as noticing and interpreting. These steps are then used to identify the time to complete tasks, and the likelihood of the user selecting the correct action in the sequence.

Summary of Pilot-Automation Interaction Models

In summarizing, we note that we did not include models of adaptive automation, because we found no efforts in which automation was adapted (rather than using automation to adapt an interface). There are many ways to model pilot-automation interaction and predict performance on the flight deck including attention and noticing changes, the design of the automation (interface, interaction), the tasks the pilot performs when using the automation, errors that the pilot can potentially commit. Because there are so many factors that can have an influence, it is difficult to capture all in a single model. The ADAT project integrates several process models applicable to the FMS in software tool. However, to date, the CASCaS effort (Lüdtke et al., 2012) appears to be the most comprehensive type of pilot performance model, addressing attention, interaction, and errors.

Conclusions

Several additional points require mentioning. One of the main findings of this review is that human performance modeling provides a viable tool for predicting pilot performance in to-be-built systems. While models typically focused on limited aspects of performance, we did note that many of the models made predictions that offered insights into potential difficulties with both existing and not-yet-developed systems. Models frequently provided useful data for comparing across conditions, and – even when predictions were incorrect – the models offered insights into pilot cognition and behavior that would have been difficult to learn otherwise. In addition, the vast majority (95%) of modeling efforts included some form of verification or validation. We believe that further efforts should be made to develop standards and guidelines for verification and (particularly) empirical validation, to support the development of more realistic and credible human performance models.

Acknowledgements

The authors would like to thank our colleagues Dave Foyle, Brian Gore, Becky Hooey, John Keller, Ronald Small, Steve Peters, Liana Algarin, and Shaun Hutchins for their contributions to this work. We also thank Tom McCloy and Michelle Yeh from the FAA for their contractual support of this effort. This work was performed under contract DTFAWA-10X-800, 05 Annex 1.11, Task Number 05-02; 09-AJP61FGI-0002.

References

Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Boehm-Davis, D.A., Holt, R.W. et al. (2002). Developing and Validating Cockpit Interventions based on Cognitive Modeling. In W.D. Gray & C.D. Schunn (Eds.) *Proceedings Twenty-Fourth Annual Conference of the Cognitive Science*, 27.

Card, S., Moran, T.P. & Newell, A. (1983). *The Psychology of Human Computer Interaction*. Lawrence Erlbaum Associates.

Gil, G., Kaber, D, et al. (2009). Modeling pilot cognitive behavior for predicting performance and workload effects of cockpit automation. *Proceedings 2009 International Symposium on Aviation Psychology*. Dayton, OH: Wright State U., 124-129.

Gonzales-Calleros, J., Vanderdonckt, J. et al., (2010). Towards Model-Based AHMI Development. *EICS '10*. June 21-23, Berlin, Germany.

Hüttig, G., Anders, G., & Tautz, A. (1999). Mode Awareness in a modern Glass Cockpit– Attention Allocation to Mode Information. Paper presented at the *10th Intl. Symposium on Aviation Psychology*, Columbus, OH.

John, B.E., Blackmon, M.H., et al.. (2009). Rapid Theory Prototyping: Example of an Aviation Task. *HFES 53rd Annual Meeting*. *53*(12), 794-798.

Lüdtke, A. & Osterloh, J-P. (2010). Modeling Memory Effects in the Operation of Advanced Flight Management Systems. *Human Computer Interaction Aero Conference 2010*, Cape Canaveral, FL.

Lüdtke, A., Osterloh, J.P., & Frische, F. (2012). Multi-criteria evaluation of aircraft cockpit systems by model-based simulation of pilot performance. *Embedded Real Time Software and Systems Conference*. Feb 1-3, Toulouse, France.

Lüdtke, A., Osterloh, J-P., Mioch, T., Rister, F., Looije, R. (2009). Cognitive Modelling of Pilot Errors and Error Recovery in Flight Management Tasks. *Proceedings of the HESSD*.

Manton, J.G., Hughes, P.K. (1990). Aircrew tasks and cognitive complexity. Paper presented at the *First Aviation Psychology Conference*, Scheveningen, Netherlands.

Miller, G. A. (1956). The magical number seven, plus or minus two. *Psychological Review* 63 (2): 81–97.

Pisanich, G.M. & Corker, K.M. (1995). A Predictive Model of Flight Crew Performance in Automated Air Traffic Control and Flight Management Operations. *International Symposium on Aviation Psychology*.

Polson, P.G., & D. Javaux (2001). A model-based analysis of why pilots do not always look at the FMA. *Proceedings of the 11th Intl Symposium on Aviation Psychology*. Columbus, OH: The Ohio State Univ.

Raeth, P.G., Reising, J.M. (1997). A model of pilot trust and dynamic workload allocation. *Proceedings of the 1997 IEEE National Aerospace and Electronics Conference* (NAECON), July 14-18.

Rasmussen, J. (1983). Skills, rules, knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics*, 13, 257-266.

Sarter, N.B., Mumaw, R., & Wickens, C.D. (2007). Pilots' Monitoring Strategies and Performance on Highly Automated Glass Cockpit Aircraft. *Human Factors*. *49*, 3. 347-357.

Schoppek, W. & Boehm-Davis, D.A. (2004). Opportunities and Challenges of Modeling User Behavior in Complex Real World Tasks. *MMI-Interaktiv*, 7, June. ISSN 1439-7854.

Sebok, A., Wickens, C., Sarter, N.et al. (2012) The Automation Design Advisor Tool (ADAT). *Human Factors and Ergonomics in Manufacturing and Service Industries*. *22*(5), 378-394.

Steelman-Allen, K., McCarley, J. & Wickens, C.D (2011) Modeling the control of attention in visual workspaces. *Human Factors*, 53, 142-153

Wickens, C.D. & McCarley, J.S. (2008). *Applied Attention Theory*. New York: CRC Press, Taylor & Francis Group.

Wickens, C.D., McCarley, J.S., Alexander, A.L., Thomas, L.C., Ambinder, M. & Zheng, S. (2008). Attention-Situation awareness model of pilot error. In D.C. Foyle & B.L. Hooey (Eds.) *Human Performance Modeling in Aviation*. CRC press.

Wickens, C.D. & Sebok, A. (2013). Flight Deck Models of Workload and Multi-Tasking: An Overview of Validation. *Proceedings of the International Symposium on Aviation Psychology*.

Wickens, C.D., Sebok, A., et al. (2013). *Modeling and Evaluating Pilot Performance in NextGen - Final Report*. FAA, Contract DTFAWA-10X-800, 05 Annex 1.11, Task Number 05-02; 09-AJP61FGI-0002.

SENSOR TO ANALYST: IMPROVED DECISION-MAKING IN AERIAL ISR THROUGH TRAINING AND DECISION SUPPORT TOOLS

Tiffany R. Poeppelman
Dan Caggiano, Ph.D.
Cullen Jackson, Ph.D.
Aptima, Inc. – Woburn, MA

Scott Galster, Ph.D.
Air Force Research Laboratory, 711 HPW/RHCP
Applied Neuroscience Branch
Wright-Patterson Air Force Base – Fairborn, OH

Capt Chad Tossell, Ph.D.
Air Force Research Laboratory, 711 HPW/RHAS
Continuous Learning Branch
Wright-Patterson Air Force Base – Fairborn, OH

Bruce Bolley
Air Force ISR Agency/A3TS
Weapon Systems Training Branch
Joint Base San Antonio – Lackland, TX

Lawrence Scarff
UTC Aerospace Systems, ISR Systems – Westford, MA

David Lind
Kevin Durkee
Sterling Wiggins
Aptima, Inc. – Woburn, MA

From sensors to analysts, modern day intelligence, surveillance and reconnaissance (ISR) requires concerted efforts from players in the aviation, aerospace, and intelligence domains to complete the Planning and Direction, Collection, Processing and Exploitation, Analysis and Production, and Dissemination (PCPAD) cycle. While the pilots, analysts, and consumers of ISR products in the PCPAD cycle have very different tasks and duties, all of them must constantly adapt to new environments, new challenges, and new enemies. The uncertainty generated by the nature of these environments places considerable decision-making and workload demands on the various operators. Currently, government, industry, and the operational community are addressing cognitive challenges placed on operators at each stage of PCPAD through training, decision aiding, and automated tools. This paper will provide a macroscopic view of the PCPAD process and discuss the current challenges being addressed within each phase.

Introduction

Good information is a critical enabler for good decisions. In its review of a series of intelligence failures leading up to the terrorist attacks of September 11, 2001, the 9/11 Commission concluded that despite readily available data indicating an attack might have been imminent, poor information analysis due to inadequate dissemination and integration of information across agencies prevented authorities from making the right decisions to prevent the attacks (National Commission on Terrorist Attacks upon the United States, 2004). Data alone is not sufficient; ". . . information is of greatest value when it contributes to or shapes the commander's decision-making by providing reasoned insight into future conditions or situations" (United States Joint Chiefs of Staff, 2011). While some efforts to improve intelligence have focused on structural changes among government agencies, far less effort has been devoted to understanding the human element in intelligence, surveillance, and reconnaissance (ISR). The Air Force's PCPAD model for ISR refers to the *planning* and direction, *collection*, *processing* and exploitation, *analysis* and production, and *dissemination* of information. In this paper, we outline the key operators and warfighters involved in PCPAD and discuss recent advances in training and decision support technologies to improve aerial ISR.

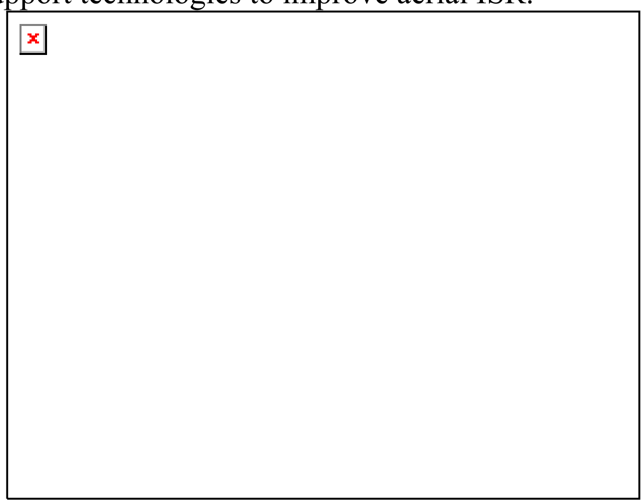


Figure 1. The PCPAD cycle involves the coordinated effort of commanding officers, dedicated field elements, and intelligence specialists to support critical decision makers at all echelons. *Photos within the image are courtesy of the U.S. Air Force.*

Who is involved in PCPAD?

Planning for an aerial ISR mission might begin when a commander at the battalion or brigade echelon identifies an intelligence gap and issues a Commander's Critical Information Requirement (CCIR). Command and operational intelligence staff may then generate priority information requirements (PIRs) from the CCIRs, select a subset of PIRs for the upcoming operation, and decompose the PIRs into Essential Elements of Information (EEIs) that specify what data is necessary to fill each PIR. Operators then *collect* information using sensors based on manned or unmanned aerial vehicles (UAVs) to address as many of the prioritized EEIs in their stack as possible during the current operation. Analysts working the sensor feed *process* and exploit what was collected to reformat the raw data and extract the information relevant to each EEI. Other analysts either co-located with the exploitation element or distributed in reach-back organizations then *analyze* patterns across the filled EEIs in conjunction with previously acquired data to re-evaluate both the PIRs and the overall understanding of the area of operations and the human terrain. To *disseminate* these analyses, the analysts must develop intelligence products, identify necessary supporting documents, and use the appropriate protocol given what is contained within the analysis before sending the products on to the appropriate consumer.

Challenges for Operators and Analysts Throughout the PCPAD Cycle

Research and development efforts in government, industry, and operational communities like the Air Force ISR Agency (AFISRA) are working to address specific challenges (cognitive, environmental, situational, etc.) placed on operators at each stage of PCPAD. The following sections will highlight how these challenges can be addressed through training, decision aiding, and automated tools and what progress has been made on these fronts.

Planning

Pre-mission planning is an on-going and essential part of successful military operations. Commanders at all echelons continually work with intelligence staff to identify gaps in the current understanding of the area of operations and of the human terrain. ISR synchronization refers to the part of the planning process in which intelligence officers: analyze Information Requirements (IRs) and intelligence gaps, identify all assets that are available, identify gaps in how the assets are being used, and make recommendations regarding which assets should collect data against the IRs and intelligence gaps (Department of the Army, 2008). ISR synchronization must consider areas, structures, capabilities, organizations, people, and events (ASCOPE) to enable commanders to make decisions based on information embedded within an appropriate socio-cultural context. While ASCOPE provides the broadest and most detailed picture possible of the area of operations and of the human terrain, managing all of the information necessary within this framework simultaneously can prove cognitively challenging for commanders charged with making critical decisions. In addition, fusing this information with the sensor assets available to collect it temporally aligned with operations is a great challenge to joining the intelligence collection planning process with the operational planning process. To address this challenge, one potential solution is to develop a tool to more seamlessly share intelligence and operational planning information across their processes and systems using novel visualizations to support decision-making during the planning cycle (Jackson, Pfautz, & Bauer, 2011).

Collection

In the collection phase, managing widely varying levels of task load is a primary challenge for pilots of UAVs. Under current concepts of operations (CONOPs), navigation and sensor operations are performed by two different crewmembers. During prolonged transits or sustained monitoring periods, workload can be quite low and lead to task disengagement (Baldwin et al, 2010; Szalma & Hancock, 2006). However, a low-workload mission can become a high-workload mission quite quickly and with little warning, such as when a potential high-value target is spotted unexpectedly. Furthermore, the increasing autonomy of UAVs has enabled future CONOPs in which a single operator performing both navigation and sensor operation functions will control multiple UAVs (Reising, 2003). Under such future CONOPS, monitoring and workload demands on the operator are only likely to increase (Parasuraman & Riley, 1997). UAV pilots currently lack a system that can adapt to these rapid fluctuations in demand and provide them with the optimal levels of control or support they currently need. On-going efforts at the AFRL seek to address this gap using neurophysiologically-based adaptive automation. Using state-of-the-art models and algorithms to integrate physiological, system, and behavioral data, AFRL's HUMAN laboratory has developed a system capable of detecting changes in operator workload in near real time (Christensen, & Estepp, in press; Pappada, Geyer, Durkee, Freeman, & Cohn, 2013). These efforts hold great promise for the future of adaptive decision aid technologies triggered by rapid detection of changes in operator state.

Processing

Processing and extraction varies in scope and duration, depending on the type of information that is being processed. Some sensors have the capability to automatically process data into a consumer-ready form, while others require manual processing and user-guided feature extraction. In the latter case, accurate processing and classification of ambiguous data can require extensive training and experience to establish fluency. Furthermore, incoming data must be monitored for time-sensitive information that must be distributed to consumers prior to extensive processing and analysis. In both cases, AFRL's Warfighter Readiness Research Division is making enormous strides in developing innovative training regimens to speed skill acquisition and retention while improving overall analyst skill level (Tossell, Jackson, Tripp, & Nelson, in press; Carlin, Jackson, Pratt, Marc, Kramer, Champagne, Kriete, & Dunlop, 2012).

Analysis

All-source and other intelligence analysts synthesize and integrate data collected from numerous sensors into meaningful patterns and predictive analyses. Analysts produce intelligence preparation of the battlefield (IPB), intelligence preparation of the operational environment (IPOE), intelligence estimates, and other products that provide leaders at all echelons of command with the information and recommendations necessary to make mission-critical decisions, yet they are plagued by two inherent difficulties in dealing with the enormous corpus of data that is collected: 'big data' management (i.e., finding the proverbial needle in a haystack when seeking a particular piece of information), and data fusion (i.e., identifying related bits of data from different sensor modalities and merging these bits together to develop a more complete picture). These difficulties have been magnified by the shift in the nature of intelligence work from what Treverton has called the 'puzzles' of the Cold War (i.e., questions with finite, knowable answers such as 'How many intercontinental ballistic missiles does the Soviet Union possess?') to the 'shapeless mysteries' (i.e., questions of speculation with no discernible answer, such as 'Will Iran develop a nuclear weapon?') of modern ISR (Moore, 2011). To address these issues, researchers are developing innovative human-computer interfaces that allow geospatial intelligence analysts to manipulate, visualize, and fuse information through a natural, multimodal interaction experience (Jackson, Pfautz, & Bauer, 2011). In addition, researchers are developing a framework for judging the value of fusing intelligence data across different sensors to determine the utility of combining the information to reduce the load on commanders while simultaneously improving their decision quality (Scarff, Jackson, Burke, Jones, Pratt, Weil, Gilfillan, & Fiore, in press).

Dissemination

Analysts disseminate the products of their work to the customers that requested the intelligence or to other intelligence consumers that may conduct forensic analyses on the newly acquired intelligence. This dissemination must be timely and efficient for the consumers to benefit from the intelligence. Given limits on communication bandwidth, particularly within forward deployed units, analysts must carefully choose with whom they share intelligence products and how the products are routed to reach the desired consumer. Redundant forwarding of information among intermediary units in the dissemination chain can tie up communications lines with the passing back and forth of the same intelligence products (Global Integrated Intelligence, Surveillance, and Reconnaissance Operations, 2012). Furthermore, intelligence products can contain supporting documents that require different security classification levels. Analysts must be sure to keep the products that are shared at the appropriate classification level for each intended consumer. While some of these concerns could be addressed with future automated, data distribution tools, decision support tools and improved training regimens would also be beneficial both for the analysts disseminating products and the consumers receiving them.

Future Directions

USAF General Norton Schwartz wrote, “We disseminate knowledge to better support decision-makers and shape operations. Still, we must never lose sight of the need for continuing to evaluate our methodologies for employing and integrating ISR capabilities vice simply increasing the density of ISR capabilities (United States Air Force, 2012).” While the operators, analysts, and consumers of ISR products in the PCPAD cycle have very different tasks and duties, all of them must constantly adapt to new environments, new challenges, and new enemies, placing considerable decision-making and workload demands on each player. The solutions outlined here are critical steps forward to mitigate these demands and will further benefit the ongoing introduction of UAVs to the civilian airspace.

Acknowledgements

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or U.S. Government.

References

Baldwin, C., Coyne, J. T., Roberts, D. M., Barrow, J. H., Cole, A., Sibley, C., & Buzzell, G. (2010). Prestimulus Alpha as a Precursor to Errors in a UAV Target Orientation Detection Task. *Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations*.

Carlin, A., Jackson, C. D., Pratt, S., Marc, Y., Kramer, D. S., Champagne, V., Kriete, R., & Dunlop, P. (2012). Contextually-relevant exploitation and analysis training environment (Unpublished technical report). Woburn, MA: Aptima, Inc.

Christensen, J.C., & Estepp, J.R. (in press). Co-adaptive aiding and automation enhance operator performance. *Human Factors*.

Headquarters, Department of the Army (2008). *ISR synchronization* (FM 2-01).

Jackson, C. D., Pfautz, S. L., & Bauer, D. (2011). Agent-driven visualizations for increasing collaboration effectiveness (Unpublished technical report). Woburn, MA: Aptima, Inc.

National Commission on Terrorist Attacks upon the United States. (2004). *The 9/11 commission report: Final report of the National Commission on Terrorist Attacks upon the United States.* Washington, DC: National. Commission on Terrorist Attacks upon the United States.

Moore, D. T. (2011). *Sensemaking: a structure for an intelligence revolution* [PDF Reader Version]. Retrieved from ni-u.edu/ni_press/pdf/Sensemaking.pdf

Pappada, S. M., Geyer, A., Durkee, K., Freeman, J., & Cohn, J. (2013, May). Modeling operational workload for adaptive aiding in Unmanned Aerial Systems (UAS) operations. Panel presented at Aerospace Medical Association Conference (AsMA), Chicago, IL.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors & Ergonomics Society*, *39*, 230-253.

Reising, J. M. (2003). *Uninhabited Military Vehicles: What Is the Role of the Operators? Proceedings of the NATO Research and Technology Organization Meeting, KN1-10.*

Scarff, L. A., Jackson, C. D., Burke, D., Jones, E., Pratt, S., Weil, S., Gilfillan, L. G., & Fiore, S. (in press). Measuring knowledge: investigative research into the quantification of performance within a contextual multi-source PED fusion process. In J. J. Braun (Ed.), *Proceedings of SPIE: Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2013 (Vol. 8756)*. Bellingham, WA: SPIE.

Szalma, J. L., & Hancock, P. A. (2006, October). Performance, workload, and stress in vigilance: The power of choice. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, No. 16, pp. 1609-1613).

Tossell, C. C., Jackson, C., Tripp, L. M. & Nelson, R. A. (in press). Optimizing data processing and management decisions during ISR through innovative training regimens. *Proceedings of the 17th International Symposium on Aviation Psychology*.

United States Air Force (2012). *Global integrated intelligence, surveillance, and reconnaissance operations* (Air Force Doctrine Document 2-0).

United States Joint Chiefs of Staff (2007). *Joint intelligence* (Joint Publication 2-0).

OPTIMIZING DATA PROCESSING AND MANAGEMENT DECISIONS DURING ISR THROUGH INNOVATIVE TRAINING REGIMENS

Capt Chad C. Tossell, Ph.D.
711th Human Performance Wing, Warfighter Readiness Research Division
Wright-Patterson AFB, OH 45433

Cullen Jackson, Ph.D.
Aptima, Inc.
Woburn, MA

Lisa M. Tripp, Ph.D.
711th Human Performance Wing, Warfighter Readiness Research Division
Wright-Patterson AFB, OH 45433

Lt Robert A. Nelson
711th Human Performance Wing, Warfighter Readiness Research Division
Wright-Patterson AFB, OH 45433

Effective intelligence, surveillance, and reconnaissance (ISR) relies heavily on both technological and human analytical capabilities. Intelligence analysts must be able to detect, interpret, process, and perform other critical tasks to turn data into meaningful information for decision-makers. The ability to aggregate massive data sets into operationally relevant information is challenging due to issues such as information overload, team coordination, time constraints, tunnel vision, and limited or vague guidance. This report describes research and development efforts to enhance training for geospatial intelligence analysts. Initial results from cognitive task analyses with these analysts along with associated technology development are discussed.

Geospatial analysts (GAs) in the United States Air Force (USAF) are responsible for planning, collecting, processing, analyzing, and disseminating (PCPAD) imagery information in order to support air, space, and cyber operations. In order to meet the high demand of the customer base, intelligence analysts must be proficient in detecting, identifying, recognizing, and correlating information to provide information critical for operational planning and execution. These analysts face a number of systemic challenges including information overload, team coordination, time constraints, tunnel vision, and limited or vague guidance (e.g., Heuer, 2005).

The current paper reports training research programs to address some of these challenges by defining characteristics of expert GAs and then developing training to optimize data processing and analysis decisions. Data processing refers to the methods employed by analysts to aggregate, correlate, interpret, and disseminate data to decision makers within the leadership chain. This communication occurs through both face-to-face and computer-supported interactions. Management decision-making refers to the interpretation and conclusions reached by analysts. These decisions can range from the identification of a specific stimulus (e.g., determining a particular building is a hospital) to a conclusion derived from synthesized intelligence information (e.g., based on multiple sources of intelligence, determining a building is clear of males). These aspects of PCPAD are critical to ensuring data is processed and turned into actionable intelligence information meaningful to decision-makers. In order to optimize these processes, we argue that training must provide relevant experiences in which trainees must leverage contextual information to accomplish their mission.

Training Geospatial Intelligence Analysts

At the broadest level, geospatial intelligence analysts (GAs) are responsible for a wide variety of tasks from analyzing the earth's geophysical structure to viewing live video feeds of an area. Largely considered the eyes of the intelligence community, GAs are responsible for interpreting the vast amounts of imagery data collected from various platforms including satellite and remotely piloted aircraft. Analysts view this imagery to accomplish tasks such as the development of patterns of life and supporting ground units. The analyst interprets these activities and creates intelligence products, typically in the form of a presentation with both imagery and textual information. This information is then used by decision-makers from senior command and control officers to ground combat personnel. Indeed, the information from GAs influences the entire air asset tasking process from initial strategy development to real-time mission execution. Thus, expertise for these intelligence analysts is critical for effectiveness.

Training Challenges

GAs face significant training challenges given the diversity in missions and areas of operation supported. This is especially true when the primary method of mission qualification training is shadowing current mission operations. While training during real missions eliminates some of the typical training challenges faced using more scholastic methods (e.g., transfer of training), it also significantly limits the diversity of training experiences to the current operational problem set. In contrast, while more scholastic methods such as classroom and standard computer-based training allow for a variety mission sets to be addressed, transfer of training to real-world missions can be a significant problem. Simulation-based training scenarios might be employed to mitigate both concerns regarding the training diversity in real-world missions and transfer of training with traditional methods. However, current simulation-based training still faces significant technological challenges when applied to the intelligence domain. Application of simulation-based training for pilots was extremely successful largely due to the type of fidelity required for effective training. The focus is often on terrain and structures (e.g., buildings) in a relatively restricted field of view. On the other hand, intelligence analysts are often tasked with viewing imagery to detect, recognize, or identify single entities within very large fields of view. The ability of current simulation technology to accurately render realistic contextual information and cues with high enough fidelity to allow for effective training is still somewhat limited. Additionally, the ability of simulation to represent the sensor data presented to analysts and allow for realistic field of view is also limited. These two issues lead to an overall lack of acceptance of simulation-based training approaches by the intelligence community at large. Future research is likely to mitigate these issues making simulation-based training for GAs more realistic.

Another significant challenge is developing training that leverages contextual information for GA tasks. Contextual information is a critical cue for the detection and recognition of suspicious activities, a tasking which is critical to effective mission support. Development of expertise in this area necessitates learning the culture and typical patterns of behavior for a particular operational area. For example, cues that might allow for insights into differentiating between males and females in certain cultures (e.g., walking next to each other) might not be the same for other cultures. Similarly, actions that might be completely normal in one context, such as a gardener digging a hole to plant tomatoes, might seem like an anomalous activity in another context (e.g., digging a hole to plant an improvised explosive device).

Although humans have a robust capacity to learn patterns, development of expertise in this type of complex pattern recognition generally requires years of extended practice (Ericsson, Krampe & Tesch-Romer, 1993). For example, work in the training literature indicates that expertise in games which rely on pattern recognition, such as chess, require years of training to reach expert levels of performance (Chase & Simon, 1973). This is a significant challenge for intelligence analysts who often work in rapidly fluctuating environments. Analysts report limited day-to-day continuity on tasking making it increasingly difficult to develop expertise in a particular area or to learn an individual's or group's typical patterns of behavior. Furthermore, these positions tend to have a very high turnover rate for multiple reasons including the military's heavy focus on career broadening opportunities, the perception of receiving lower pay than other individuals in similar career fields, and the high stress and long hours often associated with intelligence analysis work. These factors contribute to a very difficult environment to develop expertise.

Developing Training for GAs

Current training research is focused on overcoming these challenges by providing more individualized training for GAs. The approach being used to this end starts with defining expertise using a systematic process to inform adaptive training technology development. Second, the training technologies collect and track trainee experiences and provide optimal content for personalized learning based on their performance. Both of these research activities are highlighted below.

Defining GA Expertise

The first step in the construction of training methodology to facilitate and expedite the development of expertise in a domain is defining expertise within that domain. The process of defining expertise starts with understanding the characteristics that comprise the competencies, knowledge, and skills that GAs must possess in order to perform their duties and help achieve mission success. One of the challenges of understanding these characteristics is that the context of the missions may influence them—that is, the set of competencies, knowledge, and skills may vary across mission. Fortunately, a variety of analytic methods (e.g., Work Domain Analysis, Mission Essential Competency) exist to elicit these characteristics and distill them across mission areas in order to find the commonalities and differences between them. As a result, the ways in which these characteristics change with changes in context—environment, culture, equipment available, etc.—across mission areas can be known and accounted for in developing training to increase the expertise levels of GAs.

In order to understand the competencies, knowledge, and skills required to conduct geospatial intelligence analysis across a variety of mission areas, we conducted a job analysis technique called Mission Essential Competencies (Colegrove, Alliger, Beard, Bennett & Garrity, 2009; Alliger, Beard, Bennett, Colegrove & Garrity, 2007). This method involves conducting a number of workshops and administering surveys to collect data on the competencies, knowledge, and skills required for the job roles involved in the work environment, as well as information on the training environments and the types of experiences in which incumbents participate to gain expertise. In the job environment we analyzed, one of the roles was a GA. While we cannot report all of them in this paper, several key results pertain to using expertise and experience in understanding the role of context to conduct effective geospatial intelligence analysis activities. For example, understanding the operational environment—what, where, when, why—is important in clarifying tasking to ensure the appropriate data is collected. In addition, having knowledge of appropriate sociocultural factors—such as style of dress, layout of structures, and nominal behaviors—is critical to correctly process and interpret data collected. Our analysis also revealed that experts are more adept at combining knowledge across contextual factors to reveal more accurate and complete information in response to intelligence requirements; for example, combining information about observed behaviors, cultural norms, and time of day can reveal the sex of observed people.

Using Technology to Provide GA Training

As previously discussed, in order for today's GAs to gain and retain expertise in collecting, processing, and exploiting imagery data, they must be adept at utilizing contextual cues—such as terrain features, adversary activity, and typical warning signatures—in their analytical process. However, GAs face challenges that complicate their ability to gain this contextual awareness, particularly when exploiting across sensor types and facing adaptable environments and adversaries. Because each analyst will have expertise in some contexts but not others, training technology should sense and adapt to analyst competencies, knowledge and skill levels with respect to different contexts by constructing individualized training programs.

Training technology that is adaptive to the individual learner can prepare GAs to learn several contexts at once—such as target area identification, terrain analysis, and sociocultural pattern detection—by building contextual knowledge through individualized scaffolding. Adaptive training systems are computer-based training applications that utilize algorithms to determine the next learning content (e.g., event, module, course) to present and to predict the future performance "state" of the learner based on the current performance "state" of the learner. In order for these types of adaptive systems to function, the performance "state" of the learner must be measured and assessed, the learning content must be meta-tagged for what competencies, knowledge, and skills it provides for training, and algorithms must be able to reason over the meta-tagged content and the learner's performance.

One method for reasoning over these factors to provide adaptive training content is to use a statistical Bayesian approach. One such Bayesian method is called Partially Observable

Markov Decision Process (POMDP) for decision planning under uncertainty (Smallwood & Sondik, 1973). POMDP extends the classic Markov Decision Process (Puterman, 1994) and is used in diverse domains such as assisted living (Hoey, Poupart, von Bertoldi, Craig, Boutilier & Mihailidis, 2010), patient management (Hauskrecht & Fraser, 1998), and spoken dialog systems (Williams, 2010), as well intelligent training systems (Andrews, Freeman, Andre, Feeney, Carlin, Fidopiastis, & Fitzgerald, in press; Freeman, Stacy, MacMillan, Carlin & Levchuk, 2009). In fact, adaptive training systems utilizing POMDPs have been shown to reliably accelerate learning relative to a traditional strategy (hierarchical part-task training) when used to train students on a dynamic target selection task (Levchuk, Shebilske, & Freeman, 2012).

Using POMDP as part of a system to train GAs on contextual cues and information ensures that the resultant training is both adaptive and personalized since POMDP solutions continuously adjust their assessment of the student, and select the next component of the curriculum based on the student results as they are obtained. In addition, POMDP solutions or policies are really a training *plan*, which includes next and future steps of the curriculum by identifying training scenarios within the problem space through which students will gain the greatest contextual expertise given their prior performance. Finally, in order for training to most effectively transfer to the analyst's work domain, the training curricula should be designed to accommodate the analyst's typical workflow, which also would allow the system to be used for both "offline" training and mission preparation and rehearsal.

Conclusion

The goal of these efforts is to provide personalized training based on elicited knowledge from expert GAs. The key aspects to developing an adaptive training environment to support the development of expertise include: defining expert performance in terms of the competencies (knowledge, skills and experiences) critical for developing expertise, ensuring the training environment has the appropriate level of fidelity for the operational community, and integrating experiences which take into account contextual factors regarding an operational area. Personalized training will help analysts to develop the expertise to overcome challenges including information overload, team coordination, time constraints, tunnel vision, and limited or vague guidance. Future training capabilities that are being developed will leverage the above work to support and optimize data processing and management decisions across intelligence, surveillance, and reconnaissance (ISR) domains.

Acknowledgements

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, National Geospatial-Intelligence Agency, or the U.S. Government.

References

Alliger, G. M., Beard, B., Bennett, W., Colgrove, C., & Garrity, M. J. (2007). Understanding mission essential competencies as a work analysis method. AFRL Research Report

AFRL-HE-AZ-TR-2007. Mesa, AZ: U.S. Air Force Research Lab Human Effectiveness Directorate.

Andrews, D., Freeman, J., Andre, T., Feeney, J., Carlin, A., Fidopiastis, C., & Fitzgerald, P. (in press). Training Organizational Supervisors to Detect and Prevent Cyber Insider Threats: Two Approaches. *ICST Transactions of Security and Safety*.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive psychology, 4(1), 55-81.

Colegrove, C. M., Alliger, G. A., Beard, R., Bennett, W., & Garrity, M. J. (2009). M. Scott Meyers Award for Applied Research in the Workplace: Mission Essential Competencies: An operations-centric approach to improving training and job performance. Presentation at the 24th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review, 100(3)*, 363.

Freeman, J., Stacy, W., MacMillan, J., Carlin, A., & Levchuk, G. (2009, July). Capturing and building adaptive expertise in virtual worlds. In *proceedings of Human Computer Interaction International 2009*, 19-24. San Diego, CA: Author.

Hauskrecht, M. and Fraser , H. (1998). Modeling Treatment of Ischemic Heart Disease with Partially Observable Markov Decision Processes. In *proceedings of American Medical Informatics Association Annual symposium on Computer Applications in Health Care, Orlando, Florida, pp. 538-542.*

Heuer, R.J. (2005). Limits of intelligence analysis, *Orbis, 49,* 75-94.

Hoey, J., Poupart, P., von Bertoldi, A., Craig, T., Boutilier, C., & Mihailidis, A. (2010, May). Automated handwashing assistance for persons with dementia using video and a Partially Observable Markov Decision Process. *Computer Vision and Image Understanding (CVIU), 114*, 5.

Levchuk, G., Shebilske, W., and Freeman, J. (2012). A model-driven instructional strategy: The benchmarked experiential system for training (BEST). In P. J. Durlach & A. M. Lesgold (Eds.) *Adaptive Technologies for Training and Education*. Cambridge, UK: Cambridge University Press.

Puterman, M. L. (1994). *Markov decision processes: Discrete stochastic dynamic programming.* Hoboken, NJ: Wiley.

Smallwood, R. D. & Sondik, E. J. (1973). The optimal control of partially observable Markov processes over a finite horizon. *Operations Research, 21*, 1071-1088.

United States Air Force (2012). *Global integrated intelligence, surveillance, and reconnaissance operations* (Air Force Doctrine Document 2-0).

Williams, J. D. (2010, May). Spoken dialog systems as an application of POMDPs. Presentation at POMDP Practitioners Workshop: Solving real-world POMDP problems. 20th International Conference on Automated Planning and Scheduling (ICAPS), Toronto, Canada. http://users.isr.ist.utl.pt/~mtjspaan/POMDPPractitioners/

THE EFFECTIVENESS OF MICROSOFT FLIGHT SIMULATOR AS A TRAINING AID FOR PRIVATE PILOT TRAINING AND PROFICIENCY

Wendy S. Beckman
Middle Tennessee State University
Murfreesboro, TN

The debate regarding the effectiveness of computer-based software for pilot training and proficiency has been ongoing since such software first became available. While studies on the efficacy of such software have been and continue to be conducted, pilots are in large number utilizing such packages. A nationwide survey was conducted to determine how the Microsoft Flight Simulator (MFS) software package is being used by pilots for both initial private pilot training and for maintaining proficiency once certificated. Over 650 survey respondents evaluated the effectiveness of MFS in 14 areas of pilot operations. It was found that over 40% of respondents used the software package during their private training, and that 85% of respondents now use the software package to help maintain their proficiency. These findings indicate that student and private pilots have embraced MFS as a useful training and proficiency aid.

In 1980, the Microsoft Flight Simulator (MFS) software package entered the marketplace, and over the past 30 years there have been ten editions released (Grupping, 2007). The earliest versions of the software were somewhat rudimentary, as the graphics and processing capabilities of computers of the time were unable to portray flight realistically. For many years, certificated pilots did not view the software as useful for training or proficiency purposes. But now, both the software and the capabilities of even inexpensive computers have evolved so a fairly realistic flight experience is provided. This improvement has led to the use of the MFS package by pilots for both training and proficiency purposes, even though the Federal Aviation Administration (FAA) does not allow time spent using the package to be logged. There have even been two books published which describe methods for using the MFS package for flight training (Van West & Lane-Cummings, 2007; Williams, 2006).

There have been a tremendous number of studies performed to evaluate the transfer of training and effectiveness of personal computer based basic aviation training devices (BATDs) for instrument flight skills (Federal Aviation Administration, 1999; Dennis & Harris, 1998; Taylor, et al., 2003, Johnson & Steward, 2005). These studies resulted in Advisory Circular 61-136, which approved the use of BATDs for up to ten hours of initial instrument training, and for use in meeting the recency of experience requirements of 14 CFR 61.57(c)(1) (Federal Aviation Administration, 2008). However, BATDs require physical controls for the following items: landing gear, wing flaps, cowl flaps, carburetor heat control, mixture, propeller, and throttle controls. In addition, the following controls must be able to be set without using a keyboard or mouse: master/battery, magnetos, alternators, fuel boost pumps, avionics master, pitot heat, and aircraft lights. Given these requirements, MFS does not qualify as a BATD.

However, this fact has not stopped instrument pilots from using the MFS package heavily for both training and proficiency purposes. In a 2009 study, it was found that for those pilots who earned their instrument rating after 2000, approximately 80% utilized MFS during instrument training, for a mean of 51 hours (Beckman, 2009). In the same study, it was found that 61% of study participants utilize the package to maintain instrument proficiency. In both initial training and proficiency, the areas of basic attitude instrument flight, holding patterns, enroute navigation, instrument approach procedures, and previewing approaches at unfamiliar airports were ranked as skills that could be effectively practiced in MFS.

Statement of the Problem

Although instrument skills and procedures have long been practiced in simulators or training devices, the use of these training aids for private pilot training has basically been neglected. However, anecdotal evidence suggests that many student pilots not only use MFS for entertainment purposes, but also as a means to practice skills necessary to obtain their private certificate. In addition, many certificated pilots use the software package to maintain a level of proficiency when they are not able to fly regularly. Given that the necessary software and controls can be purchased for around $200, MFS potentially provides an affordable platform for both student and private pilots to enhance and retain their skills. However, if the package is to be used effectively, the skills which are most appropriate to the package need to be identified, so pilots can best use the package to their advantage. The research questions to be addressed by this study are:

1. How frequently are student pilots using the MFS package for training purposes?
2. How frequently are private pilots using the MFS package for proficiency purposes?
3. What flight skills are rated as most effective to practice with MFS during private pilot training?
4. What flight skills are rated as most effective to practice with MFS to maintain proficiency for private pilots?

Methodology

The Middle Tennessee State University Institutional Review Board approved the conduct of this human subject research study. A 12 question survey instrument was developed in electronic format using the SurveyMonkey survey generation website. The first questions concerned basic demographics, including indication of the highest certificate held by the respondent, whether they were instrument rated, whether they held a flight instructor certificate, and in what year they earned their private certificate.

After demographics, the next part of the survey was designed to see whether or not the participant used MFS during their initial private pilot training. If a participant indicated that they did use MFS during their private training, they were asked to rate the effectiveness of the software package for training in each of the following areas: General cockpit familiarization; procedures and checklists; understanding aircraft systems; interpreting flight instruments; visual cues experienced during flight; basic maneuvering flight; slow flight, stalls, and steep turns; ground reference maneuvers; pattern work and landings; VOR set up and usage; GPS set up and usage; cross country navigation; basic attitude instrument flight; instrument and equipment failures; radio procedures and phraseology; previewing unfamiliar airports; and aeronautical decision making. A standard Likert response scale was utilized, with the possible responses: Very Effective, Effective, Neutral, Not Effective, and Very Ineffective. In addition, the option to select Not Practiced was available to participants for each of the listed skills. At the conclusion of this section, participants were asked to estimate the total number of hours they spent using MFS while they were training for their private certificate.

The second section of the survey attempted to evaluate how MFS was used after the participants obtained their private certificate. Thus, the first question asked was whether or not the pilot had used the MFS package to maintain or improve their proficiency level since achieving their private certificate. If this question was answered in the affirmative, the participant was asked to indicate how effective the package was for practicing each of the following skills: Procedures and checklists; aircraft systems operations; basic maneuvering flight; slow flight, stalls, and steep turns; ground reference maneuvers; pattern work and landings; VOR set up and usage; GPS set up and usage; cross country navigation; basic attitude instrument flight; instrument and equipment failures; radio procedures and phraseology; previewing unfamiliar airports; and aeronautical decision making. The same Likert scale as indicated above was utilized for response options for this question. Participants were then asked to estimate the number of hours per month they spend using MFS to practice their flying skills, and whether they typically practiced a specific skill or fly a scenario when using MFS. Finally, participants were provided an open response area in which they could provide any additional comments about their use of the MFS package for private pilot training and proficiency.

The survey was distributed to potential participants by the daily electronic newsletter Aviation eBrief, published by the Aircraft Owners and Pilots Association. The newsletter has a subscriber list of over 415,000 pilots, aircraft owners, and other aviation professionals. Aviation eBrief published a paragraph describing the study, along with an internet link to the survey, in a November 2009 newsletter. All certificated private pilots were encouraged to participate by completing the survey. Over 745 responses were received within one week of the survey announcement, and the survey was closed to respondents at that point.

Results and Discussion

There were 745 total responses, of which 651 were valid (the respondent held at least a private certificate). Of these, 432 held only a private certificate, while 163 held a commercial certificate, and 56 held an airline transport pilot certificate (ATP). Those that did not hold at least a private certificate were not able to continue with the survey. With regards to an instrument rating, 57% of the participants indicated they had this rating, while only 18% of participants indicated having a certified flight instructor (CFI) certificate. The average year in which a private certificate was earned was 1992, with the range being from 1945 to 2009. Over 44% (281 of 629) of the respondents indicated they used MFS while they were training for their private certificate, although when the participants who had obtained their private certificate after 1980 (the year the MFS package first became available) were examined, the percentage that used it during private training increased to 61% (281 out of 459). When those that received their private certificate since 2000 were examined, it was found that 76% (213 of 282) of these participants used the package during their training.

A large majority of the 281 participants who indicated they used MFS for training responded to the follow on question regarding the effectiveness of the package for learning specific skills. The effectiveness, as perceived by the respondents, of MFS for practicing a variety of skills necessary for the achievement of a private pilot certificate, can be seen in Table 1. It was found that the six skills having a combined Very Effective and Effective rating greater than 75% included: interpreting flight instruments (97%), VOR set up and usage (91%), basic attitude instrument flight (90%), general cockpit familiarization (81%), cross country navigation (81%), and basic maneuvering flight (76%). Those skills with the lowest ratings (less than 50% of the responses in the Very Effective or Effective categories) included: ground reference maneuvers (36%), radio procedures/phraseology (42%), and aeronautical decision making (45%). All other skills were rated between 50% and 75% Very Effective or Effective. The last question in the training section of the survey asked how many hours the participant spent using MFS while training for their private certificate. The mean response for this item was 112 hours (SD = 17.62).

Table 1. *Perception of the Effectiveness of MFS for Specific Skills in Private Pilot Training*

Skill	**Very Effective**	**Effective**	**Neutral**	**Not Effective**	**Very Ineffective**	**Not Practiced**	**n**
General Cockpit Familiarization	32%	49%	11%	5%	1%	3%	261
Procedures and Checklists	19%	35%	26%	8%	3%	9%	262
Understanding Aircraft Systems	13%	38%	29%	11%	2%	8%	261
Interpreting Flight Instruments	59%	38%	3%	0%	0%	1%	263
Visual Cues in Flight	14%	37%	34%	10%	5%	2%	262
Basic Maneuvering Flight	27%	48%	14%	7%	2%	2%	262
Slow Flight, Stalls, Steep Turns	16%	38%	25%	11%	6%	4%	263
Ground Reference Maneuvers	11%	25%	29%	19%	8%	9%	262
Pattern Work/Landings	25%	36%	20%	12%	6%	2%	263
VOR Set Up and Usage	71%	20%	5%	0%	0%	4%	262
GPS Set Up and Usage	42%	26%	12%	3%	0%	16%	262
Cross Country Navigation	45%	36%	11%	4%	0%	5%	262
Basic Attitude Instrument Flight	52%	38%	6%	2%	0%	2%	262
Instrument/Equipment Failures	28%	32%	17%	5%	0%	19%	263
Radio Procedures/Phraseology	13%	29%	28%	12%	5%	13%	260
Previewing Unfamiliar Airports	37%	38%	11%	6%	1%	7%	262
Aeronautical Decision Making	13%	32%	31%	9%	2%	11%	261

The next section of the survey dealt with the use of MFS for proficiency purposes after achieving a private pilot certificate. Over 85% (n=519) of the survey respondents indicated using the package since earning their certificate. Once again, participants were asked to rank the effectiveness of the package for various skills, but this time with regards to maintaining proficiency. The results may be seen in Table 2.

Table 2. *Perception of the Effectiveness of MFS for Retaining Proficiency in Private Pilot Skills*

Skill	Very Effective	Effective	Neutral	Not Effective	Very Ineffective	Not Practiced	n
Procedures and Checklists	18%	36%	24%	5%	3%	13%	478
Aircraft Systems Operation	14%	41%	27%	6%	1%	10%	477
Basic Maneuvering Flight	23%	43%	21%	4%	2%	6%	477
Slow Flight, Stalls, Steep Turns	16%	30%	26%	11%	3%	15%	477
Ground Reference Maneuvers	9%	25%	30%	14%	5%	17%	476
Pattern Work/Landings	21%	35%	23%	10%	3%	8%	479
VOR Set Up and Usage	62%	29%	5%	1%	0%	3%	479
GPS Set Up and Usage	46%	32%	11%	2%	1%	9%	478
Cross Country Navigation	47%	35%	9%	2%	1%	5%	475
Basic Attitude Instrument Flight	61%	31%	5%	1%	1%	2%	479
Instrument/Equipment Failures	33%	36%	13%	3%	1%	15%	477
Radio Procedures/Phraseology	13%	28%	28%	12%	5%	14%	476
Previewing Unfamiliar Airports	40%	36%	11%	4%	1%	7%	481
Aeronautical Decision Making	18%	36%	27%	5%	2%	12%	474

It was found that five skills had a combined Very Effective and Effective rating greater than 75%: basic attitude instrument flight (92%), VOR set up and usage (91%), cross country navigation (82%), GPS set up and usage (78%), and previewing unfamiliar airports (76%). Those skills with the lowest ratings (less than 50% of the responses in the Very Effective or Effective categories) included: ground reference maneuvers (34%), radio procedures/phraseology (41%), and slow flight, stalls, and steep turns (46%). All other skills were rated between 50% and 75% Very Effective or Effective. Participants were then asked to estimate the hours per month that they spend using the MFS package to practice their flight skills, and the mean response was 9.6 hours (SD = .759). Finally, participants were asked to indicate whether they typically practice a specific skill or a scenario (i.e., particular flight segment) when using the MFS package. The result was 60% indicating that they practice a scenario, and 40% indicating that they practice a specific skill.

Finally, there was an open-ended qualitative question, which asked for any additional thoughts the participant had regarding the usefulness of MFS for private pilot training and proficiency. An astonishing 61% of the respondents (396 of 651) took the time to make comments regarding the package. Equally interesting, of those that responded, less than 2% had a negative comment to make regarding the use of MFS. Examples of common positive sentiments include:

- Great tool to hone your skills and keep up to date. Especially great for G1000 and other GPS procedures.
- As a private pilot, I find Microsoft Flight Simulator to be particularly effective as a tool for "pre-flying" to unfamiliar places to get the airport layout and to get the lay of the land along the route and in the airport environment.
- Excellent tool. Good to use to introduce concepts which then saves time in the cockpit.
- I got my private pilot license in 40.1 hours and passed my FAA written exam with 100% correct. I fully credit MFS for the familiarity with systems and instruments, as well as ATC phraseology.
- I use MFS to maintain basic skills, as where I live there is no general aviation light aircraft. I can get to actually fly 2-3 times a year in light airplanes. MFS is an invaluable aid in maintaining proficiency.

Again, there were very few negative comments to analyze, but the few that were present are represented by the following examples:

- Needs better visual cues for ground reference work. Turns about a point are difficult since side glances are not really effective in gauging position.

• I did not have rudder pedals, so I was weak with them.
• Can lead to bad habits such as "head inside cockpit." Need to have a clear understanding of what it is, and is not, good at to best utilize it.

Conclusions

Despite the fact that time spent using MFS cannot be credited toward a private certificate, it is clear that the package is being utilized by large numbers of student pilots. For those that received their private certificate after 1980, 61% reported using the package during training, while for those that received their private certificate after 2000, 76% reported using MFS during training. For those that used the package during training, the mean total time spent on MFS was 112 hours. The most useful training skills, as perceived by the participants, were interpreting flight instruments, VOR set up and usage, basic attitude instrument flight, general cockpit familiarization, cross country navigation, and basic maneuvering flight. The first three of these are very flight instrument and navigation instrument dependent, so it seems logical that these skills can be developed through use of a software package. The use of MFS for general cockpit familiarization also seems clear, since becoming familiar with the location of various instruments, switches, and buttons has long been done via methods such as studying a cockpit picture or diagram. Cross country navigation seems a bit less self-evident, but again, navigation instruments are often in use. While visual cues may not be as good as desired, the ability to fully practice VOR and GPS navigation, with at least some visual cues present, is a benefit. Finally, the high rank of basic maneuvering flight was a bit surprising, since again, the visual cues in the package are a bit limited. In addition, the control forces experienced in flight are quite different than those experienced in the software package. However, the ability to practice procedures repetitively while away from the airport was identified by many participants as a valuable aspect of MFS.

The 85% of participants utilizing MFS after achieving their private certificate, with a mean usage of 9.6 hours per month, is an even greater percentage than those that use the package during training. Basic attitude instrument flight, VOR set up and usage, cross country navigation, GPS set up and usage, and previewing unfamiliar airports were perceived as the most effective skills to practice with MFS. These first four skills were also indicated as highly effective in initial training as well, and the additional area of previewing unfamiliar airports has long been seen as a valuable tool by instrument pilots (Beckman, 2009). Given the difficulties of finding and entering the traffic pattern under visual flight rules at unfamiliar fields, it seems natural that the ability to examine the terrain and runway layout of a new airport would be useful to a private pilot as well.

Of the 209 participants who held their commercial or ATP certificate, 117 (56%) also held a CFI certificate. While only a small percentage of these CFIs (22%) used MFS during their initial training (perhaps because their training was conducted before the software package became as capable as it is currently), a much greater percentage (78%) of these CFIs now use the package for proficiency. Given this high rate of usage by CFIs it is likely that student pilot use of the package will continue to remain strong. This being the case, it is important for both students and instructors to know what aspects of flight training can best be assisted by MFS, and this study has attempted to bring those skills to light. In many respects, for both training and proficiency purposes MFS has replaced conventional "arm chair flying," where pilots visualized procedures and maneuvers in their head prior to flight, as a method of developing and retaining essential flight skills. Given the high cost of flight time, any cost savings or safety advantage pilots may achieve through the use of a ground training device should be encouraged, and it appears that private pilots are making wide use of MFS for this purpose.

References

Beckman, W.S. (2009). Pilot Perspective on the Use of Microsoft Flight Simulator for Instrument Training and Proficiency. *International Journal of Applied Aviation Studies, 9(2),* 171-180.

Dennis, K. and Harris, D. (1998). Computer-based Simulation as an Adjunct to Ab Initio Flight Training. *International Journal of Aviation Psychology, 8(3),* 261-276.

Federal Aviation Administration. (2008). *FAA Approval of Basic Aviation Training Devices (BATD) and Advanced Aviation Training Devices (AATD)* (AC-61-136). Washington, DC. Government Printing Office. Retrieved from http://rgl.faa.gov/Regulatory_and_Guidance_Library/rgAdvisoryCircular.nsf /0/37E40A5E69B18FEF8625748E005B327A?OpenDocument

Federal Aviation Administration. (1999). Aviation Safety: Research Supports Limited Use of Personal Computer Aviation Training Devices for Pilots. Government Accounting Office. Retrieved from http://www.gao.gov/archive/1999/rc99143.pdf

Grupping, J. (2007). Flight Simulator History. Retrieved from http://fshistory.simflight.com/fsh/timeline.htm

Johnson, D. and Stewart, J. (2005). Utility of a Personal Computer-Based Aviation Training Device for Helicopter Flight Training. *International Journal of Applied Aviation Studies, 5(2),* 287-305.

Taylor, H., Talleur, D., Bradshaw, G., Emanuel, Jr., T., Rantanen, E., Hulin, C., Lendrum, L. (2003). *Effectiveness of personal computers to meet recency of experience requirements.* (DOT/FAA/AM-03/3). Retrieved from http://www.faa.gov/library/reports/medical/oamtechreports/2000s/media/0303.pdf

Van West, J. and Lane-Cummings, K. (2007). *Microsoft Flight Simulator X for Pilots: Real-World Training.* Indianapolis, IN: Wiley Publishing, Inc.

Williams, B. (2006). *Microsoft Flight Simulator as a Training Aid*. Newcastle, WA: Aviation Supplies and Academics.

PRELIMINARY EXAMINATION OF SIMULATOR-BASED TRAINING EFFECTIVENESS

Maxine Lubner, Ph.D.
Vaughn College of Aeronautics and Technology
New York, NY
Sharon DeVivo, Ed.D.
Vaughn College of Aeronautics and Technology
New York, NY
Emerson Allen, B.S.
Vaughn College of Aeronautics and Technology
New York, NY
Andrew Dattel, Ph.D.
Marywood University
Scranton, PA.
Deb Henneberry, M.A.
Vaughn College of Aeronautics and Technology
New York, NY

A preliminary examination of the effectiveness of a simulator-based training program for pilots was conducted. Thirteen students of varying backgrounds, with limited flight experience enrolled in an intensive, simulator-based flight training program. Six students completed their FAA Private Pilot certificates in an average of five weeks, while five completed within four weeks. They completed instrument ratings within three or four weeks. These qualifications were completed with fewer flight hours than the US average. One student stopped training, but four remaining students have completed their FAA commercial, multi-engine and instructor qualifications in a timely and cost-effective manner. A combination of intensive classroom, simulator and traditional in-aircraft instruction was successful. Camaraderie and shared learning experiences were considered important to successful completion of the students' flight training. Numerous questions still surround this training approach, including implications for instructional techniques, and students' depth and continuity of learning.

Aviation simulators have been a part of flight training since the dawn of aviation. The precursor to the modern aviation simulator, the Link Trainer, was created as an efficient form of flight training that could be conducted outside of the airplane. There is overwhelming evidence that simulator based training combined with training in an airplane is an effective approach to learning (Hays, Jacobs, Prince, & Salas, 1992). In some situations, learning new aviation maneuvers in a personal computer aviation training device (PCATD) can yield better performance than when learning the maneuver in an actual airplane (Ortiz, 1994).

Simulator centric training provides several advantages: First, cost is significantly reduced when simulators are utilized. Because simulator based training is considerably less expensive than in-aircraft training, cost is reduced in any positive training effectiveness ratio comparison. Second, overall training time is reduced simply because simulator training can take place when inclement weather prohibits in-aircraft training. Third, virtually any scenario can be created in a simulator. The lessons taught by scenario applications can be performed in a slow enough manner to ensure that all the important details are taught and understood. Fourth, by freezing the simulator, issues can be discussed as they arise with full concentration, and without the resources needed to fly the airplane. Fifth, the advantages of part-task training are greater in a simulator. Part-task training allows a student to concentrate on certain aspects or elements of a scenario while placing less emphasis on other parts (Gopher, Weil, & Bareket, 1994; Wightman & Lintern, 1985). By evaluating information during the time of action, flight instructors can better assess students' conceptual understanding of situations when part-task training is implemented. A greater conceptual understanding is particularly important for complex aviation maneuvers, non-routine conditions, and situation awareness (Dattel, Durso, & Bedard, 2009). One example of part-task training is to allow students to control the aircraft while the instructor handles the task of using the throttle. Another less commonly employed example is to have the student use only the throttle while the instructor operates the other airplane controls. Performing these exercises in a simulator allows the additional and important opportunity to return to the small building blocks or steps making up those tasks while engaging the student's conceptual understanding of the procedure. Because, as in this example, the simulator records the student's actions, this allows the value-added opportunity for review and reflection of each task component by the student and the instructor. Sixth, by incorporating scenario-based training, students are able to develop mental models that permit them to hone judgment and decision-making skills for a variety of situations (FAA, 2008).

Because high fidelity simulation is not required for positive transfer of training (Taylor et al., 1999; Salas, Bowers, & Rhodenizer, 1998), capital investment in aviation simulators is becoming increasingly affordable. An important component for teaching is using conceptual versus procedural approaches, which are independent of simulator fidelity (Hawkins, 1997). Conceptual training is easily related to the discovery that scenario-based instruction is useful in incorporating decision making, as well as for mastery of certain techniques, such as how to fly in a traffic pattern. Procedural training remains important, for example in teaching landing techniques (Dattel, et al., 2009). We believe that instructors use the simulator for procedural training, often at the expense of conceptual training (Dattel, et al., 2009, Dattel, Kossuth, Sheehan, & Green, 2013). Simulator training can easily incorporate conceptual, procedural, scenario, collaborative and individual types of training.

Another neglected area related to simulator training, somewhat independent of fidelity, includes instructors' competence and acceptance of simulation-centric training . Instructors' design of the curriculum and their lesson plans using the simulator as a tool are important in producing positive results. Overall, individual instructor effectiveness (AOPA, 2010) has been found crucial to ensuring positive and satisfying pilot training. It is possible that using the simulators effectively would help to standardize and incorporate best practices in flight instruction. It is probable, however, that the increased fidelity of simulators has lulled instructors into using it as an airplane rather than developing new and more sophisticated uses related to simulation itself. Instructors themselves may not have considered the advantages of simulator training; nor have they developed sophisticated lesson plans to be used in the simulator. Instructors may not be aware of the evidence regarding conceptual versus procedural training; how to best incorporate decision making skills along with basic techniques of flying, especially when using scenarios; and employing collaborative or other techniques in a simulator. For example, most instructors find the simulator helpful for instrument, not private pilot training, probably because the simulator display more closely resembles flying an airplane in instrument conditions.

Thus, it should be cautioned that the simulator itself is not what makes training effective; rather a multi-factorial, instructional model incorporating content and procedures contributes to an optimal outcome. Related to this issue, it is clear that simulators are generally considered useful and advantageous. It is our contention that simulators are not yet used in an optimal and widespread manner. This situation probably exists because, in part, our understanding of the components of flight instruction needs to be revised in light of the advantages and restrictions of this more sophisticated training tool.

Other elements likely to be important in designing optimum flight training programs, are related to simulator centric training and scenario based training : Social and psychological components of instruction, including collaborative and individual techniques, mechanisms to advance pilots' decision making skills, and increased motivation, create useful attitudes and uncover gaps in comprehension.

Vaughn College recently embarked on a simulator-centric flight training program in partnership with an entity in the southwest region of the country - far from our New York campus - where a cohort of students traveled to the location for flight training. This arrangement allowed us to observe the effects of social bonding and collaboration on flight training achievements, including noting students' reduced anxiety and shared reward systems as discussed below. Opportunities to observe psychological components of reflective learning (Drago-Severson, El; Helsing, Kegan, Popp, Broderick, & Portnow, 2001), such as being able to rehearse, comprehend and retain knowledge were also numerous and pervasive. Similarly, there were many opportunities to teach decision making skills by following the practices of pilot experts and to spend a large amount of time following instructor-guided practice to acquire flight skills needed for safety and proficiency (Lubner, Adams, Hunter, Sindoni, & Hellman, 2003).

Methods and Program Description

Vaughn College contracted with a simulator manufacturer to offer simulator-centric flight training at their flight school in the southwest US. Students were screened before being allowed to enter this program. Students had to meet requirements of a GPA of 3.0 or better, obtain financial counseling, possess an FAA Class III Medical, take a demonstration flight, passed the FAA Private Pilot Knowledge exam and agree to remain substance-free during their training period. Preparation for the FAA Knowledge exams is incorporated into a college course that is also required for this program. With the exception of two students, the first group started their flight training with less than 20 hours that had been distributed sporadically, and one had only completed a demonstration flight.

In January 2012, the first group of eight students traveled together from New York to the mid-west training location. Students lived in a hotel and were transported to and from the flight school six long days per week for a total of six weeks. After 4 ½ weeks, because of a bad-weather week and to avoid losing time in their regular college classes, students had to return to New York. They completed their flight training for the FAA Private Pilot Certificate during spring break in March 2012.

Between January and March 2012, one student left the program after about a week of flying to pursue a career other than commercial flight, and one returned to traditional flight training in New York. In June 2012, the group of six students returned to the flight school and within five weeks, obtained their FAA Instrument Ratings.

Whenever the students were at the flight school, their training was intensive. Students were expected to study and prepare for their lessons on a daily basis. The students generally averaged six days of flying per week. As the program progressed, circumstances such as extreme mid-day summer temperatures and extended flight times during cross-country training required adjustments to the flight schedule for physiological reasons. Usually they flew twice a day and spent an average of two hours per day in the simulators. A video system was set up so that if students were practicing in the simulators on their own, any instructor could offer correction. Because this was a new program, there were inefficiencies in the beginning that appear to have improved over time.

In September 2012, the group of six returned to start training to obtain their FAA Commercial, Multi-Engine and CFI qualifications. The plan had been to obtain all these qualifications within two months, but it was decided to defer training for the CFI until January 2013. Unfortunately, one of the students passed away in October (unrelated to flight training) and precipitated the return of the entire group. This became an important learning lesson for all: Although most of those involved wanted to and felt that they could manage to continue with their flights. After a critical incident debrief and some days of reflection students were returned to New York for further grief and post traumatic stress resolution. This cohort returned to flight training after an appropriate break and with a renewed emotional connection with each other and a productive learning experience similar to their performance prior to the incident.

From mid-November through December 2012, the group of five returned to the flight school and four completed their FAA Single and Multi-Engine Commercial qualifications in what was a total of 9 weeks. The fifth person could no longer meet the flight performance requirements and has currently left the program.

The four remaining students plan to return in May 2013 for their CFI, CFI-I and ME-I for a month.

The second group of five students began their Private Pilot training in July 2012. All completed their Private Pilot Certificates within four weeks. In December 2012, these students had an additional 15 hours of simulator training at Vaughn in New York, in preparation for their instrument flight training. When they returned to the flight school in January 2013, four completed their Instrument Ratings within three weeks, while the fifth student finished in 4 weeks. This student started and ended late, and by being out of step with the timing of training, missed some of the instrumental and social group support.

A third group of five students will start their Private Pilot flight training in mid-March 2013.

Shortly, we will have created a comparison group, as two students will begin training at a school in New York. These students will use the simulators at Vaughn College and will travel together to the local flight school for at least 2 full days per week. The simulator instruction at Vaughn College will be the same for all students, whether at the remote or local flight schools. However, at the remote location, students may practice on their own in the simulators for an unlimited amount of time while they are there.

The structure of the training for the students at the remote location included:

- One flight instructor to every two students with a rotation of flight instructors after solo
- Students fly in Cessna 172s with Garmin 1000s

So far (March 2013), of the total of 9 students remaining in the first two cohorts:

- 4 who started January 2012 have achieved up to the commercial (single and multi-engine); they are expected to have added their CFI's with single-, multi-engine and instrument ratings by June 2013
- 5 who started July 2012 have achieved their private pilot certificate with instrument rating; they are expected to complete their training by the end of 2013.

For more details about the training experience, see the blog at: www.vaughn.edu

Results and Discussion

The students in this simulator-centric program were able to complete their flight training in a cost- and time-effective manner with implications important for flight training practices and research. Although the numbers are small and no good comparison group exists yet, this experience has generated testable hypotheses and speculative results.

It appears that the concentrated, simulator-centric, camaraderie-infused program has reduced attrition, reduced the costs to completion for flight training, and highlighted ways that instructors could better use simulators.

Attrition: The often used standard is that 80 percent of students who pursue a private pilot license never achieve the desired license. In Group One we initially saw six out of eight, or 75 percent, achieve their private pilot license; those six also achieved (100 percent) their instrument rating. Of those six, four (66 percent) achieved their commercial single- and multi-engine. In Group Two, five began the program and achieved their private pilot license (100 percent) and all five achieved their instrument rating (100 percent).

Time/cost to completion: Vaughn is very fortunate to be working with a partner who sets the price for a particular license or rating that is guaranteed. This means the price does not change no matter how much time the student needs. So far, students have taken anywhere from 1 week to 1 month longer to achieve their license/rating with no additional cost. This fixed cost approach does alleviate some of the long-time issues with flight training and the uncertainty around cost (AOPA, 2010). Students also have an in-depth conversation with Vaughn's financial aid office before embarking on the program to make sure that they understand and, most importantly, can afford the financial burden of flight training.

Instructors' use of simulators: We observed that instructors did not arrive with in-built knowledge of how to teach in a simulator-centric mode and that this skill was developed with each subsequent cohort. While the flight instructors perceived the simulators as teaching tools during the process of flight training—they were clearly a supplement and not necessarily used as a "first-line tool." That has begun to change as the simulators are now being truly integrated using a scenario-based approach.

Collaboration/camaraderie: As will be discussed below in terms of some of the qualitative work that has been done, the social connections made within the cohort significantly contributed to the learning (Drago-Severson et al, 2001) as the students supported and challenged one another to grow as pilots. Because this program's flight training took place far from our New York campus, the cohort of students had to train together to achieve a respective FAA certificate or rating. This requirement uncovered a crucial element – collaborative training. Because students had to go in groups to the flight school that offered simulators and airplanes in one location, they developed a strong set of social bonds.

We believe that their camaraderie afforded numerous advantages too: opportunities for reflective learning as they discussed their experiences, collaborative learning as they shared information and techniques, reduced anxiety when they encountered learning plateaus, and shared reward systems as each member of the group achieved a successful learning element. This phenomenon is documented in the literature on expertise where it is found that experts, including pilot experts, tend to have friendships with others in their fields, spend their leisure time as well as their work time in their fields and retain strong motivation to continue learning in their fields (Lubner, Adams, Hunter, Sindoni, & Hellman, 2003).

From interviews that were conducted with the first group of students and from informal discussions with both groups on an ongoing basis, several themes have emerged:

1. Keeping the same instructor until solo and then changing instructors gave the student the opportunity to learn from different teaching techniques. Students new to flight training found this type of instructor assignment helpful.
2. Having a flight-qualified adult coach/monitor on-scene became valuable as a coping strategy for the students. That person was able to help the students resolve issues with training questions, relationships with instructors and personal situations.
3. The students quickly became like a family, with all the advantages of "family" support. This was particularly useful for young students in the work-intensive environment. The support of fellow students and the presence of a flight coach greatly contributed to the ongoing reflection that was done throughout each license or rating. Reflection as a part of an experiential learning process is particularly useful when it can be managed by an educator (Fenwick, 2001), in this case our "flight coach."
4. Because of the camaraderie that developed, students were able to comfortably share some mistakes and outcomes with each other which they might not have shared with their instructors.
5. Reflecting on learning was possible in formal and informal manners. Students who were waiting at the flight school for their lessons were able to practice in simulators, watch videos, read or talk to instructors or to each other. In the hotel or over meals, students could talk and laugh about their experiences.

A more effective structure for beginning each course became apparent:

6. It was shown that if students started their flight training prior to passing the required FAA knowledge test, they fell behind while they had to split time between studying for the written and reviewing material for flight lessons. They also fell out of step with the other students, thereby lessening the collaborative advantage.
7. Before arrival at flight training it was beneficial for new students to be briefed as to what to expect. This 'scaffolding' or organizing technique reduced the "fire hose" effect in the first days of the training.
8. Students in the first group resisted practicing in the simulators because they felt that their learning was not the same as in the aircraft. While some may see this as a simulator fidelity issue, proper and effective instructor utilization of simulator-centric training methods as the program developed, did positively alter the students' perceptions.
9. While the training was intensive (ten-hour days, six-day weeks) the course improved as its structure evolved. Student performance was enhanced by the instructor adhering to detailed simulator-centric lesson plans, providing clear direction for studying, being better at prioritizing instruction, and providing relevant study guides and materials in addition to the standard reading materials. Students universally responded positively as more structure was added to the program.
10. Soft issues: The hotel used initially for most of 2012 lacked cooking facilities for the guests. While local transportation was often arranged, the arrangement (combined with the cost of eating out) led to poor eating habits. Most students did not get sufficient exercise during the time in training, and also became sleep deprived over time as they tended to stay up late studying.
11. The second student group's training was more efficient at every level. Their time-to-completion was less than Group One. Some of this was attributed to the training school's developing maturity, but a very significant reason was also the mentoring provided by Group One to Group Two. The Group One students were excellent at passing on what they had experienced, and the same is being noticed of Group Two / Group Three.

Between training sessions and back in New York, students were engaged in catching up with their academic, family and social lives and managed to use simulators at Vaughn College and went together to rent aircraft for local flights. Their motivation and enjoyment of flying remained high.

In March 2013, Groups Four and Five will begin training for their private pilot certification. Five students will be training at the distant facility used by the previous groups. Two additional students will begin training at a local flight school in the New York area. Efforts are being made for the local students to benefit from collaboration, camaraderie and simulator-centric training as did the previous groups.

It will certainly require longer-term observation and inclusion of comparison groups to carefully address questions regarding the relative importance and operation of the flight instruction components discussed above, as well as students' long-term safety retention of their proficiencies. In particular, a successful, time- and cost-limited training program raises questions about the FAA standards for initial pilot certification. For now, FAA can comfortably adhere to the current minimum flight time and proficiency standards. This is because, if the national average is 60-80 hours to attain a Private Pilot certificate (AOPA, 2010), then the FAA's minimum criterion of 35-40 hours is easy to uphold. On the other hand, if students consistently achieve and retain their initial pilot qualifications and safety standards within the minimum time required, then FAA standards for certification will need more testing and possibly, revision.

Clearly, a consistent, larger-scale flight training program that successfully limits costs and time-to-completion of initial pilot qualifications would have excellent implications for reducing the looming global pilot shortage (AOPA, 2010).

A well-designed training program using conventional and simulator-centric training, incorporating camaraderie, and instructor proficiency in this form of instruction, has early indications of being successful.

References

Aircraft Owners and Pilots Association (AOPA) (2010) *The Flight Training Experience: A Survey of Students, Pilots and Instructors*. http://www.aopa.org/

Dattel, A. R., Kossuth, L., Sheehan, C. C., & Green, H. J. (2013). *Poster presented at the 84th Annual Meeting of the Eastern Psychological Association*. New York, NY.

Dattel, A. R., Durso, F. T., & Bédard, R. (2009). Procedural or conceptual training: Which is better for teaching novice pilots landings and traffic patterns? *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society*. San Antonio, TX.

Drago-Severson, El; Helsing, D; Kegan, R.; Popp, N; Broderick, M; and Portnow, K. (2001). The Power of a Cohort and Collaborative Groups. *Focus on Basics 5, Issue B (October 2001): 15-22*. http://www.eric.ed.gov/PDFS/ED508685.pdf

FAA(2008). Aviation Instructor's Handbook. Retrieved from http://www.faa.gov/about/office_org/headquarters_/avs/offices/afs/afs600

FAA (2008) Advisory Circular *(AC) 00.2-15*. Retrieved from www.faa.gov.

Fenwick, T. (2001). *Experiential learning: A theoretical critique from five perspectives.* Columbus: Ohio State University.

Gopher, D., Weil, M., Bareket, T. (1994). Transfer of skill from a computer game trainer to flight, *Human Factors*, 36, 387-405.

Hawkins, F.H. (1997). *Human Factors in Flight.* (2nd ed.). Brookfield, VT: Ashgate.

Hays, R. T., Jacobs, J. W., Prince, C., Salas, E. (1992). Flight simulator training effectiveness: A meta-analysis, *Military Psychology*, 4, 63-74.

Lubner, M., Adams, R., Hunter, D., Sindoni, R. and Hellman, F. (2003). Risks for aviation accidents or incidents among U.S. pilots by pilot training, experience and exposure. *Twelfth International Symposium on Aviation Psychology*. Dayton, OH.

Ortiz, G. A. (1994). Effectiveness of PC-based flight simulation. *The International Journal of Aviation Psychology,* 4, 285-291.

Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology*, 8, 197-208.

Taylor, H. L., Lintern, G., Hulin, C. L., Talleur, D. A., Emanuel, T. W., & Phillips, S. I. (1999). Transfer of training effectiveness of a personal computer aviation training device. *The International Journal of Aviation Psychology*, 9, 319-335.

Wightman, D. C., & Lavern, G. (1985). Part-task training for tracking and manual control. *Human Factors*, 27, 267-283.

Paper to be presented at: 17th International Symposium on Aviation Psychology, May 6 – May 9, 2013, Wright State University, Dayton, Ohio.

THE FAA'S HUMAN FACTORS AIR TRAFFIC CONTROL / TECHNICAL OPERATIONS STRATEGIC RESEARCH PLAN

Edward Austrian
Fort Hill Group
Washington, DC USA

Dino Piccione
Federal Aviation Administration
Washington, DC USA

The Federal Aviation Administration (FAA) Air Traffic Control (ATC) / Technical Operations (TO) Human Factors Team completed development and publication of a ten-year strategic research plan. This effort was endorsed by the FAA Research Engineering and Development Advisory Committee (REDAC) and sponsored by the FAA Human Factors Division. The intent of the strategic plan is to define a clear pathway for the ATC / TO Human Factors Team to improve and integrate human factors research in ATC applications. The strategic plan also serves as a communication tool to engage the community of practice and promote the implementation of products that yield the greatest operational results—which will support the achievement of the Team's mission as well as the FAA's transition to a future, automated National Airspace System (NAS).

The FAA is engaged in an extensive effort to modernize the NAS. This phased, incremental transition involves the introduction of new ATC automation, technologies, and procedures to replace legacy equipment and services. The modernization to an automated NAS aims to achieve the system-wide benefits of increased safety, capacity, and efficiency. However, "understanding the human factors work that is being or needs to be done" to transition to an automated NAS "is a fundamental challenge to the practice of program management in research and development" (Krois et al., 2010**)**. In response to this challenge, the Air Traffic Control (ATC) / Technical Operations (TO) Human Factors Team developed the ATC / TO Human Factors Strategic Research Plan. The Plan provides the ATC / TO Human Factors Team and community of practice with a clear, ten years strategic direction, and measurable success criteria that support the transition to an automated NAS. The strategic plan was developed in alignment with Destination 2025, the FAA Advanced Concepts and Technology Development Directorate mission, and operational needs identified by internal FAA sponsoring organizations. Execution of the strategic plan supports achievement of the ATC / TO Human Factors Team's mission (FAA, 2012a):

> "Provide human factors research, engineering, and leadership to enhance operational human performance in the NAS. Our findings and products will enable the FAA to achieve its mission to provide the safest, most efficient airspace system in the world"

Strategic Plan Content

Strategic plan content was defined through the execution of a requirements analysis to determine the current state of human factors in the NAS, a high level gap analysis, and the identification of opportunities for future research advancements based on industry needs and Agency level direction. The strategic plan is composed of five research categories that are defined by related objectives and lower level strategies. A summary of document categories, objectives, and strategies are listed below in Table 1. The remainder of this document will detail strategic materials generated as a result of the aforementioned activities.

Table 1.
Research Categories, Objectives, and Strategies

Category: Research for Operations	
Objective 1: Provide In-Service / Post Implementation support to the AT and TO Organizations to enhance fielded systems, procedures, and operations	**Strategy 1.1:** Develop a Change Management strategy for the ATO to transition from current operations to an automated future NAS
	Strategy 1.2: Integrate HF into day-to-day operations to improve human performance and safety in the NAS
Category: Human Centered Design	
Objective 2: Implement human centered design concepts to achieve operational human performance objectives, reduce the likelihood of human error, and increase the availability of AT systems	**Strategy 2.1:** Develop an integrated ATM philosophy for service-oriented workstations
	Strategy 2.2: Develop HF system standards and guidelines to improve human-system performance
Category: Human Systems Integration	
Objective 3: Lead the development and incorporation of the human systems integration concept into an automated, future NAS	**Strategy 3.1:** Integrate applied human factors into the development of a human centered automation philosophy
	Strategy 3.2: Further develop the Air-Ground Integration concept to improve controller-pilot situational awareness and decision making abilities
Category: Selection and Training Process Improvement	
Objective 4: Strengthen the FAA's personnel selection and training process to improve the hiring and maintaining of a qualified, diverse, workforce of the future	**Strategy 4.1:** Improve existing AT selection tests to reduce Agency costs of hiring ATCSs through refined selection
	Strategy 4.2: Improve applicant placement by facility type to increase training efficiencies and decrease developmental attrition rates
	Strategy 4.3: Expand technical and personal development training to improve skills and abilities of the workforce to perform job functions and maintain the safety of the NAS
	Strategy 4.4: Improve Technical Operations workforce selection, training, and workforce maintenance
	Strategy 4.5: Proactively assess AT and TO job-tasks to identify emerging KSAOs during the transition to an automated NAS for future selection process improvement
Category: Increase Human Performance and Safety	
Objective 5: Maintain a high level of human performance and safety within the current and future NAS as new technologies and operations are introduced to the AT system	**Strategy 5.1:** Improve NAS actor performance so the probability of human error within the AT system is reduced
	Strategy 5.2: Support the use of Safety Management Systems to proactively identify, manage, and mitigate human factors risk(s) to the NAS and identify contributing AT incident and accident causal factors to promote the future optimization of ATC and TO performance
	Strategy 5.3: Improve AT system availability through the reduction of unplanned AT system outages as a result of human error

Research for Operations

One of the most difficult challenges faced by the FAA is meeting the expectations and operational needs of all system users. Simultaneously, safety, capacity, efficiency, and predictability must be increased during operations in a seamless global environment (FAA, 2012b). To address these forward looking challenges, the ATC / TO Human Factors Team has made it a priority to engage in all segments of the FAA Acquisition Management System (AMS) process, with an emphasis on the In-Service Management phase. Within this phase, service organizations are responsible and accountable for operational sustainment activities, corrective actions, and management of personnel as well as technical resources. Numerous research opportunities exist to positively impact ATC / TO systems, operational human performance, safety, coordinated decision-making, and human factors policy inputs.

A future automated NAS will introduce changes to the roles and responsibilities of NAS actors across the various domains as well as the allocation of functions between human operators and automation (Krois et al., 2010). Responding to change, whether it is on a human or system level, may be challenging if systems or procedures are not well presented and received by the user-community. In an effort to streamline the change process from idea to implementation within the ATO, the ATC / TO Human Factors team will engage in change management research. The centralized theme of change management products will be human factors processes to transition individuals, teams, and organizations from a current state to a desired future state. R&D products are intended to assist the ATO in proactively maximizing system and procedural user-acceptance while also reducing potential safety risks.

Successful transition to an automated NAS is dependent on newly implemented systems and procedures meeting post-implementation success criteria that incorporate the end-user. Historically, FAA human factors research has focused on assessing replacement system performance against legacy system performance with limited facility personnel interactions. To better engage the operational community, improve workplace conditions, facility culture, safety, training, and inclusion in the FAA AMS process, the ATC / TO Human Factors Team will conduct exploratory research to integrate human factors personnel on a day-to-day basis at the facility level. Due to the complexity, size, diversity, and number of facilities across the United States NAS, management of human factors resources to accomplish and sustain this strategic objective is one of the primary challenges facing the ATC / TO Human Factors Team.

Human-Centered Design

“A major challenge is integrating human factors” with future “technology and procedures to ensure safety” (FAA, 2012b). Future ATC / TO workstation design must include human factors inputs throughout the design process to achieve operational human performance objectives and reduce the likelihood of unintended operator outcomes. Historically, minimal guidance regarding the definition of human roles and responsibilities, interface development, or the incorporation of emerging technologies in FAA documentation has led to human-centered design shortfalls. To mitigate potential system, human, and automation challenges, the ATC / TO Human Factors Team will partner with industry and other government agencies to develop human factors standards and guidelines to be used as inputs to legacy replacement system acquisition, as well as updates to existing technologies. Uniformity across workstations is intended to improve human performance, system usability, and user efficiencies.

From an air traffic perspective, service-oriented workstations are currently defined via the services provided by air traffic controllers to NAS consumers in one of the three ATC domains: Tower, Terminal Radar Approach Control (TRACON), and En Route. Services provided are constrained to the limitations of facility-specific and domain-specific hardware and software, airspace configuration(s), and the inter-operability of air traffic systems with adjacent facilities. The two radar environments (TRACON and En Route) have similar functions and capabilities but operate on different automation platforms that are being independently modernized. Research opportunities exist to develop an integrated Air Traffic Management philosophy to address the application of successful human performance design elements from one domain to another to optimize human performance in future hardware and software releases. Research philosophy products are multi-faceted. In this specific case, the philosophy is intended to integrate the human-centered design and human systems integration (HSI) concepts into ATC system design. The philosophy will assist operational personnel achieve system level human performance objectives and system developers to integrate independent automation tools.

Human Systems Integration

"Human systems Integration intends to ensure the effective and efficient interaction between aviation actors and other system elements in the NAS to enhance safety and performance" **(**Krois, et al., 2010). Specific to the ATC / TO domain, factors comprising HSI include: Human Factors Engineering, Safety, Training, and Personnel Selection. Each of the four factors build off of the human centered design concept and support the identification of inter-related research needs to yield comprehensive, higher quality products for operations.

As the FAA transitions to an automated NAS, the Agency is moving from a cognitive-based aviation control system to one with automated decision support tools (DST) requiring a collaborative work environment (FAA, 2012b). Future ATC system design may increase levels of automation and DSTs. Automation will be instrumental in increasing safety and capacity only if the operator can effectively utilize the automation and DSTs. Therefore, the ATC / TO Human Factors Team will engage in research that supports the development of a human-centered automation philosophy. The automation philosophy will be oriented by the operation of domain-specific systems utilized by air traffic controllers. This perspective will promote human centered system design and act as a barrier for unintended designer errors leading to unintended operator outcomes. Additionally, the philosophy will provide standardized methods of proactively testing automation for human performance risks prior to implementation. As new automation and technologies are introduced, it is critical for there to be a formal definition of human roles and responsibilities, as well as human-automation interactions as NAS actors and automation continue to evolve to meet future needs.

Selection and Training Process Improvement

The ATC / TO Human Factors Team is responsible for proactively identifying changes in the roles and responsibilities of air traffic controllers and system maintainers that lead to updates in workforce selection and training processes. The FAA is delivering "aviation access through innovation" by "ensuring that airport and airspace capacity are more efficient, predictable, cost-effective, and matched to public needs" by having "the right people with the right skills in the right positions at the right time" (FAA, 2012b).

As the FAA transitions to an automated NAS, the Agency needs to build upon past achievements and continue to strengthen its personnel selection and training processes, while maintaining methods of selection that are valid, reliable, legally defensible, and fair. Benefits to refined selection, such as decreased costs, attrition rates, and improved training efficiencies, can be further realized. To advance the research and development of products to improve selection, the ATC / TO Human Factors Team will engage in the following activities:

- Improve applicant placement by facility type to increase training efficiencies and decrease developmental attrition rates
- Expand technical and personal development training to improve skills and abilities of the workforce to perform job functions and maintain the safety of the NAS
- Improve ATC and TO workforce selection, training, and workforce maintenance
- Proactively assess AT and TO job-tasks to identify emerging Knowledge, Skills, Abilities, and Other Characteristics (KSAOs) during the transition to an automated NAS for future selection process improvement

Increase Human Performance and Safety

Maintaining a high level of human performance and safety as new technologies, procedures, and policies are introduced to the current and future NAS is a priority to ATC / TO Human Factors Team. This will be accomplished through the prioritization of research that improves safety event data collection, development of new analytical methods to enable a better understanding of safety event causal factors, derivation of human factors interventions to prevent undesired events, and the proactive detection of emerging human error hazards. Additionally, human factors support is needed to support the FAA’s continued growth of the Safety Management System (SMS). Integration of human factors into SMS will assist in streamlining the SMS process, reduce the likelihood of isolated ATC safety decisions, and provide the opportunity to incorporate human performance hazard assessments to mitigate risks associated with procedures, technologies, and policies.

Conclusion

The ATC / TO Human Factors Strategic Research Plan is the ATC / TO Human Factors Team’s commitment to provide the FAA, its operational workforce, and flying public with the highest quality human factors research products that yield the greatest operational results. Achievement of the ATC / TO Human Factors Team mission is dependent on the effective execution of the strategic plan. Therefore, the plan is intended to serve as a communication tool for the Team to further engage the human factors community of practice in its ten year strategic direction and crosscutting accomplishments.

Acknowledgements

We would like to acknowledge the FAA’s Human Factors Division (ANG-C1) for funding this effort. Additionally, we would like to acknowledge the air traffic control and human factors subject matter experts who helped contribute to the definition and development of the strategic plan’s content.

Disclaimer

The opinions expressed are those of the authors and do not represent the Federal Aviation Administration (FAA).

References

FAA. (2012a). *Air Traffic Control / Technical Operations Human Factors Strategic Research Plan.* Retrieved from https://www.hf.faa.gov/hfportalnew/Uploads/gcreighton/ATC%20TO%20HF%20Strategic%20Plan%20November%202012%20Version%201.0.pdf

FAA. (2012b). *Destination 2025*. Retrieved from http://www.faa.gov/about/plans_reports/media/Destination2025.pdf

Krois, P., Herschler, D., Hewitt, G., McCloy, T., & Piccione, D. (2010). Human Factors Research and Development Planning for NextGen. In the *Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society*. 2010, San Francisco, CA.

EFFECTS OF NEXTGEN CONCEPTS FOR SEPARATION ASSURANCE AND INTERVAL MANAGEMENT ON ATCo SITUATION AWARENESS

Thomas Z. Strybel, Kim-Phuong L. Vu, Dan Chiappe, Corey A. Morgan
California State University, Long Beach
Long Beach, CA 90840
Vernol Battiste
San Jose State Foundation and NASA Ames Research Center
Moffett Field, CA 94350

We examined the effects of responsibility for interval management and separation assurance on ATCo situation awareness, workload and performance. Four conditions were tested by crossing two strategies for interval management (ATCo responsible or flight deck responsible) and two strategies for separation assurance (ATCo primary or automation primary). Situation awareness and workload were assessed with an online probe technique. Workload was lowest when both functions were automated, but situation awareness for conflicts depended on the sector. Both workload and situation awareness were related to the number of ATCo-managed LOS.

An ongoing challenge for The Next Generation Air Transportation System (NextGen) is to determine the optimal function allocation strategy for air traffic management (ATM) tasks. Many proposed function allocation strategies have off-loaded ATM tasks to automation or the flight deck in order to reduce air traffic controllers' ATCos) cognitive workload because airspace capacity is often limited by the current, ground-based human-centered separation assurance system. Although greater reliance on automation and/or flight deck capabilities are expected to improve the efficiency of the NAS in NextGen, the optimal allocation strategies to achieve safe operations have yet to be determined (Kopardekar et al., 2011).

Two important air traffic management functions currently performed by ATCos are separation assurance and interval management. Separation assurance refers to ensuring legal safe distances between aircraft. Interval management is maintaining a time based interval between a lead and following aircraft to achieve a time interval (required distance) over a point in space such as the final approach fix (FAF). Both functions are currently the responsibility of air traffic control (ATC), and one or both could be allocated either to ground-based automated tools or to appropriately equipped flight decks, in order to reduce ATCo workload. Research over the past 10 years (e.g., Prevot et al., 2012) has evaluated allocation strategies for separation assurance and interval management in a series of human-in-the-loop simulations. However, these projects have focused on system outcomes (i.e., efficiency and safety) and workload, but not operator situation awareness, a known determinant of operational errors (e.g., Gronland et al., 1998). Moreover, in simulations where situation awareness was assessed, most measured the construct with custom rating scales of unknown validity, making comparisons of the changes in awareness difficult to determine.

Previous research on separation assurance has examined the impact of allocating responsibility for separation assurance either to ground-based automation or the flight deck. For example, Prevot et al. (2012) showed that ground-based automation tools could be used to increase ATC performance and reduce their workload for high-density traffic sectors. For flight-deck separation assurance, responsibility is transferred to pilots. Research has generally supported the feasibility of the concept, depending on the adequacy of support tools, both on the ground and in the air (e.g., Vu et al., 2012; Strybel et al., 2013). However, allocating responsibility for separation assurance to the flight deck does not necessarily produce equivalent reductions in the tasks performed by ATCos, because the controller is typically required to intervene in exceptional situations. In a review of flight-deck-based spacing, Zingale and Willems (2009) found that self-spacing aircraft maintained more precise spacing intervals, required less vectoring and fewer air-ground communications than aircraft that are subject to controller issue spacing (existing ground-based procedures). However, Boursier et al., 2006 found that the benefits of self-spacing depend on the percentage of aircraft engaged in self-spacing.

Recently, we evaluated the impact of alternative concepts of operations for separation assurance on pilot situation awareness, workload and performance (Strybel et al., 2013). In this simulation, responsibility for separation assurance was assigned to pilots, ATCos, or an automated, ground-based agent. Using an online probe tool (SPAM; Durso & Dattel, 2004), we showed that when pilots were responsible for separation assurance, their situation awareness increased. Strybel et al. also categorized the probe queries, and showed that the increase in pilot situation awareness was specifically higher for real or potential traffic conflicts. Moreover, the increase in awareness for traffic conflicts was not at the cost of lower awareness of other flight-relevant information such as command-communications and aircraft/airspace status. The present investigation is a follow up to Strybel et al. (2013). Here, we evaluated the impact of different concepts for separation assurance and interval management, on ATCo situation awareness, workload and performance. Using the same sectors and similar concepts of operation to those in Strybel et al., we report on preliminary data regarding the effect of different function allocation of separation assurance and interval management on ATCo situation awareness, workload and performance.

Method

Participants

Fourteen retired radar-certified ATCos (9 Center, 5 TRACON) with 11 or more years of radar experience participated. All had participated in previous simulations with the simulation software used here.

Simulation Configuration

The simulation was run using the Multi Aircraft Control System (MACS), Aeronautical Datalink and Radar Simulator (ADRS), and VoiceIP software, developed by the Airspace Operations and Flight Deck Display Laboratories at NASA Ames Research Center (e.g., Prevot et al., 2012; Johnson et al., 2005). ATCs used MACS configured as a Digital System Replacement (DSR) display with integrated Data Comm and conflict alerting. Additional tools such as conflict probes, and spacing information were provided in some concepts of operation. All aircraft were flown by pseudopilots and they also managed spacing in the Automation Primary Concept for Interval Management.

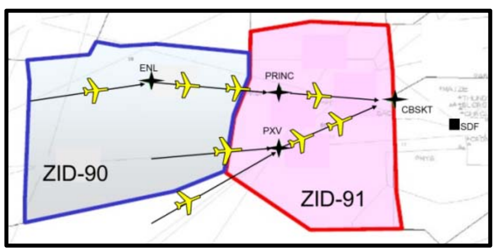


***Figure 1*. Airspace Sectors and Arrival Streams**

Scenarios

Two sectors ZID-90 and ZID-91 were simulated, as shown in Figure 1. Although sector 90 is actually a ZKC sector, it was renamed to be consistent with the simulation configuration (participants were seated next to one another in the same room). Each scenario contained arrivals to and departures from Louisville Airport (SDF), and over flights. The arrival stream consisted solely of UPS aircraft flying constant descent approaches (CDAs) on the CBSKT arrival. These aircraft entered ZID-90 from the west or southwest, and were assigned spacing lead and time en trail (105 s) upon entry. Each scenario also contained static weather cells located in the eastern portion of ZID-90. Two flight equipage types were included in each scenario, Autonomous Flight Rules (AFR) and Instrument Flight Rules (IFR). AFR flights had airborne conflict alerting and resolutions tools, and were therefore capable of self-separating. AFR flights were also permitted to deviate for weather without ATCo clearance. UPS arrivals were always on AFR flight rules. IFR flights were managed by ATCo. They were required to request weather deviations from ATCo, and had no on board conflict resolution tools.

Depending on the concept, the ATCos were responsible for separation assurance and/or interval management (spacing). In addition, the ZID-90 ATCo was responsible for handling pilot requests for weather deviations, and the ATCo in ZID-91 was responsible for managing the merging of the Pocket City (PXV) and Centralia (ENL) arrival streams out of ZID-90 to the CBSKT intersection for Louisville (SDF). Controllers in both sectors were required to respond to UPS requests for re-sequencing and spacing due to weather maneuvers. The ATCo in ZID-91 also managed SDF departures on westbound routes.

Concepts of Operation

Responsibility for resolving AFR-IFR conflicts and maintaining spacing depended on the operating concepts that are shown in Table 1. Four plausible NextGen operating concepts were evaluated, based on combinations of responsibility for separation assurance (ATC primary vs. Automation Primary) and interval management (ATC primary vs. Automation primary). For separation assurance, when ATCo was primarily responsible, all IFR-IFR and IFR-AFR conflicts were resolved by ATCo, using the trial planner and/or conflict resolution tool. When Automation was primarily responsible, AFR-AFR and IFR-AFR conflicts were resolved by the auto-resolver agent without ATCo approval. For interval management, when ATCo was primarily responsible, the UPS arrival sequence was managed by ATCo, using speed and other commands (e.g. vectoring) for spacing. When automation was primarily responsible, UPS arrival aircraft received spacing commands upon entering ZID-90, and pseudopilots engaged onboard spacing tools for interval management.

Table 1.
Concepts of operation: Separation assurance and spacing was delegated to ATCos or automation.

		Responsibility for Spacing	
		ATC Primary	Automation/Flight Deck Primary
Respon-sibility for Separation Assurance	ATC Primary	**Conflicts** ATC: IFR-IFR and IFR-AFR Automation: AFR-AFR **Spacing** ATC Initiates and maintains	**Conflicts** ATC: IFR-IFR and IFR-AFR Automation: AFR-AFR **Spacing** Automation initiates, flight deck maintains
	Auto-mation Primary	**Conflicts** ATC: IFR-IFR Automation: AFR-AFR and IFR-AFR **Spacing** ATC Initiates and maintains	**Conflicts** TC: IFR-IFR Automation: AFR-AFR and IFR-AFR **Spacing** Automation initiates, flight deck maintains

The DSR was equipped with advanced traffic management tools depending on operating concept. A conflict probe detected conflicts 8 minutes prior to a loss of separation (LOS) and alerted ATCo by flashing the conflicting aircraft pair in red and displaying the time to LOS next to the call signs. This tool was available for all flights. A trial planner allowed ATCos to manually create conflict-free flight plan changes for separation assurance and interval management, by clicking on the flight path, creating a new waypoint, and moving the waypoint and its corresponding flight path to a new location or clicking on the altitude in the data tag and selecting a new altitude. This tool was available for both AFR and IFR flights. An auto-resolver tool provided resolutions to conflicts when requested by ATCo. Both pilots and ATCos can request resolutions from the auto-resolver tool, depending on operating concept. The algorithm was not aware of weather, though. The auto-resolver agent detected and resolved conflicts without approval from ATCo. The auto resolver agent handled AFR-AFR conflicts under all concepts of operation, and AFR-IFR conflicts in the Automation-Primary Concept for Separation Assurance.

Situation Awareness & Workload Measurement.

To assess the impact of operating concepts on situation awareness and workload, we used the SPAM online probe technique. SPAM present SA probe questions regarding the operator's work environment while the simulation is ongoing and the displays active. Both accuracy and latency of responses to these queries are used as measures of situation awareness. Questions were delivered beginning four minutes into the scenario, and every three minutes afterwards. The queries began with a ready prompt on the probe display; ATCo responded as soon as he/she was able to respond to a query without disrupting his/her sector. If a response was not given within one minute, the ready prompt "timed out" and was removed from the screen. A new ready prompt was then presented 2 minutes later (preserving our 3-minute probe interval). Immediately after the ready prompt was responded to, a probe question was presented. These probes asked operators questions about the airspace and were formatted as either true/false or multiple choice questions. ATCo selected their answer to the question by touching the button corresponding to the answer on a touch-screen display. The probe questions were categorized based on the task information queried and whether the information was relevant to the ATCo's own sector or the sector of the adjacent ATCo.

The categories queried were conflicts, spacing status and traffic/weather. Conflict questions asked about potential conflicts [e.g., "In the next 3 minutes (20 miles), will there be any conflicts between an AFR- IFR aircraft pair if no further action is taken?"], and recent conflict resolutions, either by ATCo or by the auto-resolver (e.g., "Was the resolution to the last conflict an altitude change?"). Spacing-status questions asked about UPS arrival aircraft and their status regarding spacing [e.g., "Will any AC need re-sequencing in your sector in the next 3 minutes (20 miles). Traffic/weather questions asked about current status of aircraft in a sector such as number of AFR and IFR aircraft, altitudes and headings [e.g., "In the next 3 minutes (20 miles), from which direction will the majority of AFR aircraft enter your sector?"]. Questions also asked about weather: "How close is [AC callsign] to the nearest weather cell?"]. Workload was assessed with several measures, two of which will be reported here. We measured the latency to the ready prompt for probe queries because this latency has been shown to be related to workload. Workload ratings also were queried four times in each scenario.

Procedure

Each participant was tested over a one week period. On the first day, participants were assigned to one of the two adjacent sectors (ZID 90 or ZID 91) and were trained on the traffic flows and operating concepts. Data collection occurred on days 2-5, concluding with a debriefing session on the last day. The same pair of ATCos worked the adjacent sectors throughout the week. On each data collection day, one 50-minute scenario was run under each operating concept, with the order counterbalanced.

Results

In this paper, we present preliminary results from the simulation, focusing on sector safety, workload and situation awareness metrics. We computed the numbers of LOS for each concept that were the responsibility of ATCo (IFR-IFR LOS in automation primary, IFR-IFR + IFR-AFR LOS in ATC primary). The mean number of ATCo-responsible LOS was 1.1 (SEM=.13) in ZID-90 and 1.3 (SEM=.18) in ZID-91. Workload was measured by the ready latencies, the time between presenting the ready query and the ATCo's acceptance of it, and by the responses to the workload queries themselves. Situation Awareness was assessed in terms of response latency and response accuracy to situation awareness probe questions for each category of question.

The numbers of LOS were significantly correlated with the ready response latencies ($r = .18$, $p = .02$) such that longer response latencies were associated with greater numbers of LOS. The mean workload rating was also correlated with LOS ($r = .42$, $p <.001$), with higher workload ratings associated with more LOS. Ready latencies were also correlated with workload ratings ($r =.28$; $p <.001$). Both workload measures indicate that higher workload produced more LOS. However, both measures were also significantly correlated with probe latencies (r's =.17 and -.18 for ready latency and workload ratings, respectively). Therefore, to determine the effectiveness of situation awareness probe queries in predicting sector safety, we computed partial correlations between probe latency and probe accuracy after removing variance due to workload. Probe latency, when collapsed across information categories, was significantly and positively correlated with number of ATCo-responsible LOS (*pr =.17, p < .05).* Probe latencies for spacing (*pr =.16, p < .05)* and traffic/weather queries (*pr* =.18, $p < .05$) were also significantly correlated with LOS. Interestingly, latencies for conflict questions were not significantly related to number of LOS. Partial correlations between probe accuracy and number LOS were not significant except for accuracy on traffic/weather probes, which was marginally significant.

To determine the effect of operating concept on workload, mixed ANOVAs were run on ready latencies and workload ratings, with repeated-measures factors Separation Assurance Concept and Spacing concept, and the between-subjects factor, Sector. For ready latencies, log transformed values were used, due to violations of normality. For both ready latencies and workload ratings, significant main effects of separation-assurance concept were obtained [ready latency: $F(1,12) = 46.96$; $p < .001$; workload rating: $F(1,12) = 98.07$; $p <.001$], as well as significant interactions between separation-assurance and spacing concepts (ready latency: $F(1,12) = 8.47$; $p = .02$; workload rating: $F(1,12) = 5.22$; $p = .05$), as shown in Figure 2. For both measures, workload was lower in the automation-primary separation assurance concept than in the ATC-primary separation concept. For automation-primary separation, the lowest workload was obtained when spacing was also automated. Note, however, that when ATCo was primarily responsible for separation assurance, no difference was obtained for spacing concepts based on ready latency, and the difference between workload ratings for spacing concepts is much smaller than in the automation primary-separation concept. This suggests automating conflict resolution produces greater reductions in ATCo workload compared to when automation is responsible interval management.

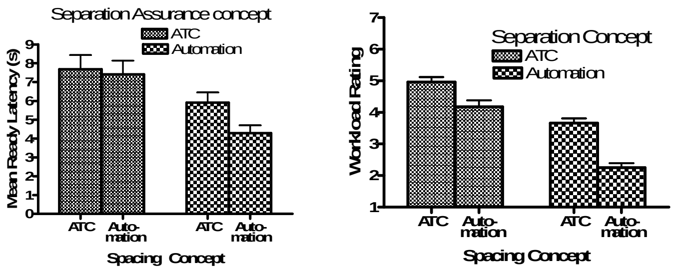


Figure 2**. (Left) Mean ready latency and (Right) Mean Workload Rating for Each Separation and Spacing Concept.**

Situation awareness probe latencies were analyzed similarly. Probe latencies were evaluated because accuracy was not related to sector safety. Mixed ANOVAs were run on the mean probe latency for each question category: conflicts, spacing and traffic/weather. For conflict probes, a main effect of separation-assurance concept ($F(1,12) = 7.91$;$p = .02$) and marginal main effects of spacing concept ($p = .08$) and sector ($p = .06$) were obtained. A three-way interaction between these variables also was significant, $F(1,12) = 7.18$; $p = .02$, as shown in Figure 3.

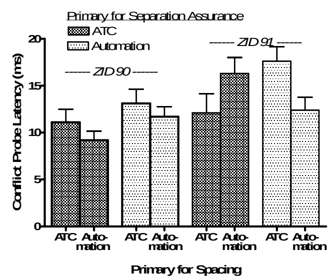


Figure 3**. Three-way interaction between sector, separation concept and spacing concept on conflict probe latency.**

Based on conflict probe latencies, controllers in ZID-90 (*M* = 11.4 s, *SEM* = 6.3 s) were more aware of conflicts than controllers in ZID-91 (*M* = 14.6 s; *SEM* = 8.2 s). In ZID-90, latencies for the ATC-primary separation assurance concept were lower overall than for the automation-primary concept, suggesting that ATCos had higher conflict awareness when they were responsible for detecting and resolving most of the conflicts. Moreover, greater awareness of conflicts was observed in ZID-90 with automation-managed spacing regardless of separation-assurance concept. In fact, the highest conflict awareness occurred for ATC-managed separation and automation-managed spacing. In ZID-91 a different picture emerged. First, the simple effect of separation assurance concept was nonsignficant, because when automation was primary for separation assurance, greater awareness of conflicts was found with automation-primary spacing, similar to ZID-90. However, when ATC was responsible for separation assurance, greater awareness was observed when ATC was primary for spacing. In effect, in ZID-91, greater conflict awareness occurred when both functions were either automated or controlled by ATCo.

Discussion

Our measures of workload and situation awareness were related to the number of ATCo-managed LOS. When automation was primary for separation assurance, only IFR-IFR LOS was the responsibility of the controller. When ATCo was primary, IFR-IFR and IFR-AFR LOS were the responsibility of the controller. Although the number of LOS obtained in the present simulation are higher than what is typically expected, it is important to keep in mind that this level of LOS is not surprising given the high traffic density we used in the simulation. We found that both workload and situation awareness metrics were related to the number of LOS.

The results from the workload data are straightforward: workload was lowest when automation was responsible for both spacing and separation assurance. However, when only one function was automated, automated responsibility for separation assurance produced lower workload than automated responsibility for spacing. This was not surprising considering that ATCos had to monitor more aircraft, including those AFR aircraft that might be a threat to an IFR aircraft, when they were primarily responsible for separation assurance. For the ATCo-primary spacing conditions, the ATCos were responsible for fewer aircraft.

However, the cost of lower workload with automated separation assurance and spacing may lie in lower ATCo situation awareness, with awareness for some information more affected than others. Awareness of conflicts was highest when ATCo was primary for separation assurance and automation primary for spacing. This may be due to the lower workload experienced for the automation-primary spacing condition when ATCo was responsible for separation. This finding was limited to controllers in ZID-90, in which most of the spacing was initiated. In ZID-91, awareness of conflicts was highest (probe latencies lowest) when both concepts were either automated or managed by ATCo. Assuming that fewer spacing clearances were required because most AC were sufficiently lined up in the upstream sector, the ZID-91 ATCo may have been more aware when both concepts were consistent.

References

Boursier, L., Favennec, B., Hoffman, E., Rognin, L., Trzmiel, A., Vergne, F., et al. (2006a). Airborne spacing in the terminal area: Controller experiments on mixed equipage,abnormal situations, and transition. EEC Note No. 24/06, Brétigny-sur-Orge, France.

DiMeo, K., Kopardekar, P., Ashfod, R., Lozito, S., & Mackintosh, M-A. (2001).Shared-separation: Empirical results and theoretical implications. *The Fourth International Air Traffic Management R&D Seminar ATM-2001*, Santa Fe, NM.

Gronlund, S.D., Ohrt, D., Dougherty, M., Perry, J.L. & Manning C.A. (1998). Aircraft Importance and its Relevance to Situation Awareness. Report DOT/FAA/AM-98/16. Federal Aviation Administration.

Johnson, W.W., Battiste, V., Granada, S., Johnson, N.H., Dao, A. Q., Wong, D. & Tang A.B. (2005) A simulation evaluation of a human-centered approach to flight deck procedures and automation for en route free maneuvering, *Proceedings of the 12th International Symposium on Aviation Psychology*, Oklahoma City.

Kopardekar, P.H., Aquilina, R., Edwards, T.A. Crisp, V.K. (2011). Next Generation Air Transportation System, Concepts and Technology Development Project: FY2011-2015 Project Plan (Version 3.0). National Aeronautics and Space Administration.

Prevot, T., Homola, J., R., Martin, L.H., Mercer, J.S.,& Cabrall, C.D. (2012). Towards automated air traffic control – investigating a fundamental paradigm shift in human/systems integration, *International Journal of Human-Computer Interaction*, 77-98

Strybel, T.Z., Vu, K.P. L., Battiste, V. & Johnson, W. (2013). Measuring the impact of NextGen operating concepts for separation assurance on pilot situation awareness and workload. *International Journal of Aviation Psychology*,

Vu, K.-P.L., & Strybel, T.Z., Battiste, V., Lachter, J., Dao, A-Q.,V., Brandt, S., Ligda, S. & Johnson, W. (2012b). Pilot performance in trajectory-based operations under concepts of operation that vary separation responsibility across pilots, air traffic controllers, and automation. *International Journal of Human-Computer Interaction*, 28, 107-118.

Zingale, C. & Willems, B. (2009). Review of aircraft self-spacing concepts: implications for controller display requirements. Federal Aviation Administration, DOT/FAA/TC-TN-09/03.

Acknowledgements

This project was supported by NASA cooperative agreement NNX09AU66A, *Group 5 University Research Center: Center for Human Factors in Advanced Aeronautics Technologies* (Brenda Collins, Technical Monitor).

PERCEIVED USEFULNESS OF PLANNED NEXTGEN CAPABILITIES BY AIR TRAFFIC CONTROL TOWER CONTROLLERS

Jon Holbrook
San Jose State University
Moffett Field, CA

Bonny Parke
San Jose State University
Moffett Field, CA

Ray Oyung
San Jose State University
Moffett Field, CA

Rebecca Collins
San Jose State University
Moffett Field, CA

Kari M. Gonter
San Jose State University
Moffett Field, CA

Bettina L. Beard
NASA Ames Research Center
Moffett Field, CA

The Next Generation Air Transportation System (NextGen) is an ambitious collaborative effort between government agencies and industry to increase the capabilities of the current air traffic system. Under NextGen, tools are being developed to support air traffic controllers in many aspects of their jobs. However, budgetary constraints, unanticipated technological hurdles and other challenges to implementation make it unlikely that every NextGen tool in development will find its way into future air traffic control facilities. Information is needed to prioritize NextGen tool development to ensure that the tools that will provide the most benefit can be implemented in the appropriate facilities. Toward this end, a web-based survey was conducted of 174 air traffic control tower controllers to identify the perceived usefulness of 10 planned NextGen tower capabilities designed to provide or support: departure metering at the ramp, taxi routing, departure runway assignments, departure flow management, runway scheduling, runway configuration management, integrated arrival/departure scheduling, enhanced surveillance, electronic flight data, and tower data communications. Along with brief descriptions of each of the planned capabilities, surveyed tower controllers were asked to indicate the extent to which each capability would help them in their job and affect capacity, efficiency, flexibility, predictability and safety at their airport. Results indicated that different NextGen capabilities were perceived as useful across different tower facilities. Implications for the prioritization of NextGen tool development are discussed.

The FAA's Next Generation Air Transportation System (NextGen) is a transformational plan to enhance air traffic operations through the introduction of advanced automation. The FAA has identified numerous capabilities that it plans to automate for Air Traffic Control Tower (ATCT) operations by 2018. Computational simulations of these capabilities suggest that some of the metrics (operational efficiency, capacity, flexibility, safety, predictability and environmental enhancements) used to show NextGen success can be achieved theoretically. However, most of the algorithms under development are still too immature to be tested or have not been tested with humans in the loop (Beard, Johnston and Holbrook, 2013). Our overarching goal is to develop an additional metric, a measure of human performance, that must be measured before FAA stakeholders can claim success.

Although air traffic controllers and traffic management coordinators will be the primary users of the new automation, their participation in NextGen decisions has been limited (GAO, 2009). Previous National Airspace System (NAS) modernization efforts have shown that insufficient input and buy-in by the users of new systems can delay certification and at times result in unintended application of the new systems. The approach taken here was to obtain ratings from ATCT controllers of how helpful they think NextGen improvements would be to their job (the human performance estimate) as well as to operational capacity, efficiency, safety, flexibility and predictability. To reach a large sample size and broad range of facility types, we distributed a survey that could be accessed and taken over the internet. Based on other research at NASA (Beard, Parke, Holbrook, & Oyung, in preparation; Holbrook, Puentes, Stasio, Jobe, McDonnell & Beard, 2011) we hypothesized that controllers would give lower ratings to suggested improvements to automate their more complex cognitive tasks and decisions and higher ratings to those improvements that would enhance their awareness of the situation.

Method

Participants

Survey responses were received from 174 tower air traffic controllers. Data from 126 controllers were included in this analysis, because data from the remaining 48 respondents were incomplete. The included participants represented 52 tower facilities, including 16 Core 30 airports (i.e., airports identified by the FAA as those with significant activity serving major metropolitan areas and/or as hubs for airline operations). Participants from Core 30 airports represented 42% (n=53) of survey respondents, with the remaining 58% (n=73) of respondents coming from non-Core 30 airports. The tower facilities represented by the participants ranged in complexity across FAA facility levels 5 through 12.

All study participants were current tower air traffic controllers, 89% of whom were Certified Professional Controllers (CPCs), and the remaining 11% were CPCs in training, or developmentals. Participants varied in terms of their years of experience in air traffic control: 36 participants had less than 10 years of experience, 25 participants had 10-19 years of experience, and 65 participants had 20 or more years of experience.

Materials

Although a primary goal of this survey was to inform policy makers and technology developers about controllers' views on NextGen, the complexity of NextGen created a challenge for developing a survey that active-duty controllers could take and complete in their spare time. The FAA has proposed 82 technological or procedural changes for implementation in the 2015-2018 timeframe in the form of 31 Operational Improvements (OIs) and 51 associated OI-increments (https://nasea.faa.gov). To provide a more concise and manageable description of NextGen plans, our research team distilled these plans into the following 10 capabilities, which were described in terms of tools to support tower controllers' tasks,:

1. **Departure metering at the ramp** – This tool will suggest gate release or taxi times for individual aircraft to help meet scheduled departure times and/or traffic restrictions.
2. **Taxi routing** – This tool will suggest taxi routes for individual aircraft to help organize the physical queue of traffic in the movement area.
3. **Departure runway assignment** – This tool will suggest runway assignments for departure aircraft to help distribute traffic across available runways.
4. **Departure flow management** – This tool will suggest takeoff times and opening/closing of departure routes or fixes to help align with terminal and en route airspace Traffic Management Initiatives.
5. **Runway scheduling** – This tool will suggest queue-release sequences for departure aircraft, and landing or runway-crossing sequences for arriving aircraft to help organize the flow of traffic on runways.
6. **Runway configuration management** – This tool will suggest which runways should be used for arrivals and departures to help distribute traffic across available runways.
7. **Integrated arrival/departure scheduling** – This tool will suggest schedule and staging information for arrivals and departures to help provide dynamic flexibility for managing traffic.
8. **Enhanced surveillance** – This capability will track and identify vehicle and aircraft positions, as well as provide monitoring and alerting for runway incursions and taxi conformance through both cockpit and tower displays.
9. **Electronic flight data** – Electronic flight data will replace analog data (including paper flight strips) to provide immediate access to and sharing of up-to-date flight-plan and aircraft status information among controllers, airlines and automated tools.
10. **Tower data communication** – This tool will supplement voice communication by providing the ability to transmit automated terminal information, departure clearances and amendments, and taxi route instructions electronically between aircraft and controllers.

We also identified several metrics for evaluating these NextGen capabilities based on key operational metrics that NextGen was designed to enhance, as well as controllers' individual performance, including:

- **Your job** – Helping you carry out your work responsibilities
- **Airport capacity** – Maximizing the number of operations safely conducted in a given time period
- **Airport efficiency** – Minimizing delay in gate-to-gate operations
- **Airport safety** – Identifying and mitigating loss of separation and aircraft confliction risks
- **Flexible operations** – Adjusting operations as needed in real time

- **Predictable operations** – Providing consistent and dependable information to support planning

Participants for the survey were recruited through coordination with a NATCA representative, who vetted the survey and sent a link to the survey website to NATCA members by email.

Procedure

Upon visiting the survey website, participants were provided with a brief overview of the goals and content of the survey. They were reminded that their participation in the survey was voluntary, that the survey was anonymous, and that their data would be protected in accordance with NASA's privacy policy. No incentives or compensation were provided for participation.

Survey participants were asked to identify the perceived usefulness of the 10 planned NextGen tower capabilities described above. For each capability, controllers were asked to use a sliding scale from -5 to +5 to indicate the extent to which each capability would impact their job and affect capacity, efficiency, safety, flexibility and predictability at their airport. Controllers could view brief descriptions of each of the planned capabilities, as well as definitions of the operational performance metrics, by hovering over the capability or metric name with their cursor.

In addition to rating the NextGen capabilities, controllers were given an opportunity to comment on improvements they would like to see in the tools and procedures they use today, as well as suggest new tools or procedures that would assist them in their job. Controllers also provided demographic data, including their current position, years they have worked in ATC, their 3-letter airport identifier and facility level.

Results

Controllers submitted 174 surveys, of which 126 were complete. Only data from the 126 completed surveys were included in the data analysis. Participants completed the survey in a median of 10 minutes, with a range from 4 to 279 minutes.

For each capability (e.g., departure metering at the ramp) and for each metric (e.g., airport capacity), t-tests were performed to determine if the mean value for that capability or operational metric differed from zero in either direction. For all statistical tests, significance was set at a = .05. T-test results are shown in Table 1.

Table 1.
T-test results for Capability and Metric to determine if mean ratings differed from zero.

	Mean (S.E.)	Statistical Test
Capability		
1. Departure metering at the ramp	1.01* (.15)	t(125) = 6.87, p < .001
2. Taxi routing	0.14 (.17)	t(125) = 0.80, p = .424
3. Departure runway assignment	0.23 (.17)	t(125) = 1.32, p = .189
4. Departure flow management	0.56* (.14)	t(125) = 3.84, p < .001
5. Runway scheduling	0.14 (.17)	t(125) = 0.81, p = .420
6. Runway configuration management	0.38* (.17)	t(125) = 2.23, p = .028
7. Integrated arrival/departure scheduling	0.75* (.16)	t(125) = 4.63, p < .001
8. Enhanced surveillance	1.60* (.13)	t(125) = 11.85, p < .001
9. Electronic flight data	1.31* (.15)	t(125) = 8.75, p < .001
10. Tower data communication	1.04* (.19)	t(125) = 5.48, p < .001
Metric		
Job	1.09* (.14)	t(125) = 7.67, p < .001
Capacity	0.73* (.12)	t(125) = 5.92, p < .001
Efficiency	0.97* (.13)	t(125) = 7.27, p < .001
Safety	0.74* (.12)	t(125) = 6.42, p < .001
Flexibility	0.02 (.13)	t(125) = 0.15, p = .880
Predictability	0.75* (.13)	t(125) = 5.95, p < .001

Note: * denotes significance at a = .05; S.E. = Standard Error

Survey data were also subjected to an analysis of variance (ANOVA) with within-subjects factors of Capability (10 levels, see above) and Metric (6 levels: job, capacity, efficiency, safety, flexibility predictability) and between-subjects factors of Experience (3 levels: 0-9 years [n = 36], 10-19 years [n = 25], and 20+ years [n = 65]) and Facility (4 levels: small non-hub [n = 48], medium non-hub [n = 25], medium hub [n = 27], and large hub [n = 26]). Levels of the Facility variable were determined by combining airport identifier and facility level data in the following way: non-Core 30 airports with facility levels 5-7 were identified as small non-hubs (e.g., GTF, TOL, FSM); non-Core 30 airports with facility levels 8-9 were identified as medium non-hubs (e.g., PDX, LGB, DAB); Core 30 airports with facility levels 8-10 were identified as medium hubs (e.g., IAD, MEM, PHX); and Core 30 airports with facility levels 11-12 were identified as large hubs (e.g., LAX, CLT, ORD).

Analyses revealed a main effect of Capability, $F(9,1026) = 12.83$, $p < .001$. Within-subjects contrasts revealed that participants rated Capabilities 1, 8, 9, and 10 higher than the overall mean rating, and Capabilities 2, 3, and 5 lower than the overall mean. A main effect of Metric, $F(5,570) = 30.76$, $p < .001$, was also detected. Within-subjects contrasts indicated that participants rated Job and Efficiency higher than the overall mean, and Flexibility lower than the overall mean. Analyses also showed a marginal main effect of Experience, $F(2, 114) = 2.30$, $p = .105$. Planned comparisons revealed that the mean rating by controllers with 20+ years (0.46) was lower than the mean rating by the controllers with 0-9 years of experience (1.03), $p = .043$. Mean ratings for controllers with 10-19 years of experience (0.92) did not differ from the other groups.

A significant interaction of Capability by Facility, $F(27,1026) = 1.63$, $p = .022$, along with subsequent post-hoc tests, indicated that controllers from large hub facilities rated the impact of Capabilities 8, 9 and 10 lower than did controllers from small non-hub and medium hub facilities (see Figure 1).

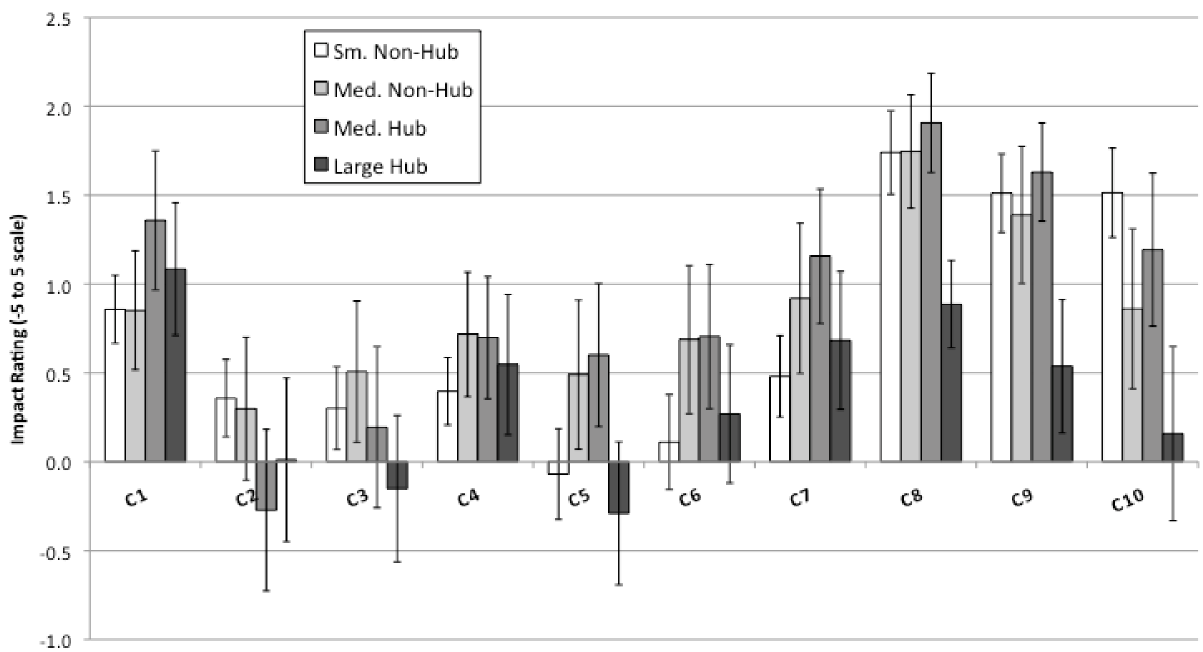


Figure 1. The interaction between Capability and Facility type indicated that controllers from large hubs gave lower impact ratings to Capabilities 8, 9 and 10. Error bars represent standard error.

A significant interaction of Capability by Metric, $F(45,5130) = 21.97$, $p < .001$, and follow-up post-hoc tests indicated that controllers rated the impacts of the operational performance metrics differently across Capabilities. This interaction is shown in Figure 2.

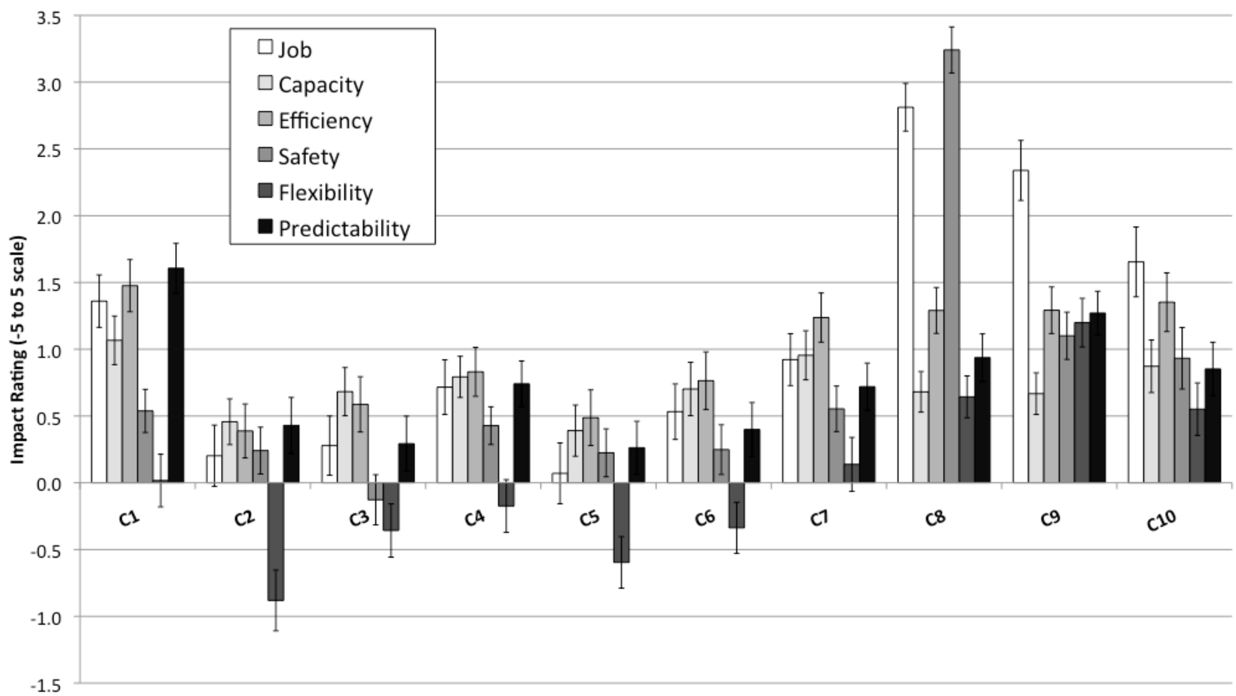


Figure 2. The interaction between Capability and Metric. Error bars represent standard error.

Significant differences across metrics within a Capability are depicted in Table 2. For each Capability, impact ratings for metrics highlighted in light gray were rated higher than those highlighted in dark gray. Metrics shown with a white (i.e., not highlighted) background did not differ from the highlighted groups.

Table 2.
Significant differences across metrics within each Capability are shown using light/dark gray highlighting.

Capability	Metric				
1. Dep. metering	Job	Capacity	Efficiency	Predictability	Safety	Flexibility
2. Taxi routing	Job	Capacity	Efficiency	Predictability	Safety	Flexibility
3. Runway assign.	Job	Capacity	Efficiency	Predictability	Safety	Flexibility
4. Flow mgmt.	Job	Capacity	Efficiency	Predictability	Safety	Flexibility
5. Runway sched.	Job	Capacity	Efficiency	Predictability	Safety	Flexibility
6. Runway config.	Job	Capacity	Efficiency	Predictability	Safety	Flexibility
7. Integ. arr./dep.	Job	Capacity	Efficiency	Predictability	Safety	Flexibility
8. Surveillance	Job	Capacity	Efficiency	Predictability	Safety	Flexibility
9. Elec. flt. data	Job	Capacity	Efficiency	Predictability	Safety	Flexibility
10. Data comm.	Job	Capacity	Efficiency	Predictability	Safety	Flexibility

Note: Metrics highlighted in light gray were rated higher than those highlighted in dark gray at a = .05.

Discussion

The 10 NextGen capabilities described above can be grouped into three categories: Capabilities that support tactical decision making by controllers (i.e., 1. Metering at the ramp, 2. Taxi routing, 3. Departure runway assignment[1], and 5. Runway scheduling); capabilities that support strategic decision making by traffic managers (i.e., 4. Departure flow management, 6. Runway configuration management, and 7. Integrated arrival/departure scheduling); and capabilities that support situation awareness (i.e., 8. Enhanced surveillance, 9. Electronic flight data, and 10. Tower data communication).

[1] Departure runway assignment is initially a traffic management task, but assignments can be changed by the controller based on the immediate tactical situation. Additionally, at smaller facilities, this function might be performed primarily by a controller.

Overall, the results of this study supported our hypothesis that controllers would rate most highly the capabilities that support situation awareness, and rate lowest the capabilities providing tactical decision making (i.e., those supporting complex cognitive tasks). Prior research indicates that controllers are disrupted and lose situational awareness when automation performs their tactical decision tasks (e.g., Holbrook, Hoang, Malik, Gupta, Montoya, & Jung, 2012). The capabilities providing strategic decision making primarily support traffic managers, and would have less direct impact on tower controllers. Controllers' high ratings on the Job metric for capabilities 8, 9 and 10 are consistent with the expected impact of tools that would enhance a controller's situational awareness. Interestingly, controllers at large hub facilities rated the capabilities that support situation awareness lower than did controllers from small- or medium-sized facilities. However, large hub controllers were also more likely to express concerns about heads-down versus out-the-window time in their comments, for example: "Controllers are already looking out the window less than they used to. New automation could make this worse." This issue of heads-down time is a concern for situation awareness tools that are tied to tower displays (e.g., ASDE-X, electronic flight strips).

Controllers gave relatively high ratings to Capability 1, departure metering at the ramp, even though it falls into the category of tactical decision support (and therefore has a greater potential to disrupt planning and controller situation awareness). This concern was perhaps acknowledged by controllers through a lower safety rating relative to the other metrics for that capability. However, Capability 1 received the highest ratings of any capability on the capacity, efficiency and predictability metrics, indicating that tower controllers see this as an area for high potential impact, particularly at airports with a higher volume of scheduled operations, as suggested by the higher ratings by hub over non-hub controllers. As one controller from a small non-hub facility stated, "I can see how these ideas will benefit the busier airports, but are not applicable to us due to our lower traffic volume."

Controllers were fairly consistent in giving low impact ratings for the Flexibility metric. Controllers also voiced this concern in their comments, for example: "Airlines have a hard time hitting a 10-minute EDCT window. The more exact we try to make it, the worse the outcome will be. There has to be a fudge factor for the system to work." This issue of flight crew compliance is critical for the success of NextGen algorithms, which rely heavily on aircraft meeting 4-D trajectories. The same capabilities (i.e., 8, 9 and 10) that support controller situational awareness should also support flight crew situational awareness, increase flight crew compliance with 4-D trajectories, and thus enable more successful implementation of the other NextGen capabilities.

This study helps to address a concern raised by the GAO (2009) that active air traffic controllers have not been appropriately involved in NextGen planning, and provides a glimpse at a potential indicator of the impact of NextGen enhancements on human performance.

Acknowledgements

We would like to thank NATCA NextGen representative Mel Davis for his support in survey development and distribution. This work was funded under Interagency Agreement #DTFAWA-09-X-80020 between NASA and the FAA. The FAA Human Factors Research & Engineering Group coordinated the research requirement and its principal representative acquired, funded, and technically coordinated execution of the research service.

References

Beard, B. L., Johnston, J. C., & Holbrook, J. (2013). NextGen operational improvements: Will they improve human performance? *Proceedings of the 17th International Symposium on Aviation Psychology.* Dayton, OH: The Wright State University.

Beard, B. L., Parke, B., Holbrook, J., & Oyung, R. (2013). Survey of human factors experts on the potential human performance risks and benefits of NextGen capabilities. Manuscript in preparation.

Government Accountability Office. (2009). *Next generation air transportation system: Issues associated with midterm implementation of capabilities and full system transformation*. GAO-09-481T.

Holbrook, J., Hoang, T., Malik, W., Gupta, G., Montoya, J., & Jung, Y. (2012). Reducing environmental impact while maximizing airport throughput: The consequences of introducing new operational goals with and without automation support in an air traffic control tower simulation. *Proceedings of the 4th International Conference on Applied Human Factors and Ergonomics,* 1819-1828. San Francisco, CA: USA Publishing.

Holbrook, J., Puentes, A., Stasio, N., Jobe, K., McDonnell, L., & Beard, B. L. (2011). How thoroughly do proposed NextGen mid-term operational improvements address existing threats? *Proceedings of the 16th International Symposium on Aviation Psychology.* Dayton, OH: The Wright State University.

DISPLAY OVERLOAD: AN ARTIFACT-BASED WORK ANALYSIS OF AIR TRAFFIC MANAGEMENT

Alicia Borgman Fernandes
Chris Brinton
Mosaic ATM, Inc.
Leesburg, VA

Air traffic management personnel manage traffic flows into, out of, and through the area for which air traffic controllers in their facilities are responsible. We report on a cross-domain, artifact-based, work analysis of air traffic management that employed a hierarchical modeling methodology centered on the artifacts used in the operational setting and the information and decision-making support they provided. Many tools support traffic managers in planning for, managing, and monitoring traffic flows, but information and decision support functions are rarely integrated across tools. Thus, traffic managers spend much of their time acquiring and assimilating information from multiple displays and manually entering information into multiple systems. NextGen initiatives such as System-Wide Information Management, En Route Automation Modernization, and the Terminal Flight Data Manager represent opportunities to implement a data exchange standard that facilitates display and decision support tool integration.

Air traffic management personnel in Traffic Management Units (TMUs) manage traffic flows into, out of, and through the geographic area for which their facilities are responsible with the goal of using the National Airspace System (NAS) as efficiently as possible (Federal Aviation Administration, 2010). An artifact-based approach was used to analyze the environment of TMU personnel in three different domains: Air Route Traffic Control Centers (ARTCCs), Terminal Radar Approach Controls (TRACONs), and Air Traffic Control Towers (ATCTs).[1]

The analysis focused on the tools used by TMU personnel and the key tasks accomplished using those artifacts. Those of most interest are coordination tasks, many of which involve acquiring, assimilating, and sharing information across domains via the TMU. These tasks require traffic management personnel to acquire information from several tools and share it across domains using several different tools. This information sharing often must be done manually, creating opportunities to introduce errors into communication and reporting.

Some manual data entry tasks could be automated by improved integration of TMU tools. However, many tools are developed by different organizations for different purposes. Emerging technologies can be leveraged to create data exchange standards to support integration of TMU tools and functions and decrease the data entry burden on traffic management personnel.

This paper provides a brief background of air traffic management and task analysis methods. It also describes the analysis performed in this effort and key results, and proposes a method for supporting TMU integration.

---

[1] Few ATCTs have formal TMUs. Some have Traffic Management Coordinators, while all have personnel that coordinate with other facilities and perform surface management functions.

Air Traffic Management

To carry out their responsibility to maximize NAS efficiency, traffic management personnel perform functions such as (Federal Aviation Administration, 2008):

- Monitor the status of NAS resources, often including constraints some distance away.
- Anticipate constraints that will impact the ability of aircraft to operate efficiently in airspace under the control of their facility.
- Develop programs such as Traffic Management Initiatives (TMIs) to modify traffic flows and maintain efficiency despite the constraints.
- Coordinate with traffic management personnel in other facilities to implement TMIs, including neighboring facilities and facilities in other air traffic control domains.
- Coordinate with flight operator personnel as necessary to maintain traffic flows.
- Report events of interest to other air traffic control facilities

Traffic management personnel use a variety of tools to accomplish these goals. Some tools facilitate traffic management planning, providing information such as weather and traffic demand forecasts. Others allow personnel to record traffic management activities and share NAS information within the facility and with other facilities to facilitate cross-domain coordination. However, few of these tools are integrated with each other, requiring mental information fusion and manual entry of data into multiple systems.

The lack of integration of TMU tools is a well-known but rarely documented issue. Borgman and Smith (2010) reported that several TMU personnel interviewed noted the lack of integration and duplicative data entry requirements as issues they would like to see resolved in new traffic management systems. Similarly, Nadler (2005) discussed ways in which the National Traffic Management Log (NTML) could be used to decrease manual data entry requirements for TMU personnel. However, in both cases the need for improved integration of TMU tools was only a tangential theme, and the issues cited have not been addressed.

Artifact-Based Work Analysis

A variety of work analysis methods have been documented and used to understand work as performed and the environment(s) in which it is performed. See Annett (2004) for a brief history of task analysis methods. Some methods make use of the artifacts of the domain to identify how they support cognitive work and to guide the design of new technologies to support joint cognitive systems (Hutchins, 1995; Nemeth, Cook, O'Connor, & Klock, 2004; Woods & Hollnagel, 2006). Artifacts often represent boundaries between system components and can provide insights into ways in which systems can be improved.

This work used artifacts of the air traffic management domain to identify cross-domain coordination processes and ways in which practitioners achieve this coordination using the tools provided to them. The analysis is discussed in the next section.

Artifact-Based TMU Analysis

The artifact-based analysis identified the tools at each TMU position in an ARTCC, TRACON, and ATCT, the users of each tool, the work process(es) that used each tool, and the frequency with which the tool was used. Due to space limitations, discussion is limited to the identification of tools and the work processes they support.

Identify Tools at Each TMU Position

The first step in the artifact-based analysis was to visit ARTCC, TRACON, and ATCT facilities and document the tools and displays at each TMU workstation. An ARTCC TMU, for example, includes several positions such as Supervisory Traffic Management Coordinator (STMC), Coordinator, En Route Spacing, Departures, and Severe Weather. When a TMU is fully staffed, there is at least one person working at each station. However, a TMU is rarely fully staffed (because of both lack of need and lack of staffing availability) and therefore each person working in the TMU often must work multiple positions.

Figure 1 shows the Supervisory Traffic Management Coordinator (STMC) position at one ARTCC. The STMC has several responsibilities (FAA, 2010; FAA, 2008) related to managing the air traffic operation, coordinating with other facilities and other fields of expertise such as Technical Operations, keeping data accurate in a number of air traffic management systems, and responding to “special situations that may arise” (FAA, 2008, p. 17-4-1).

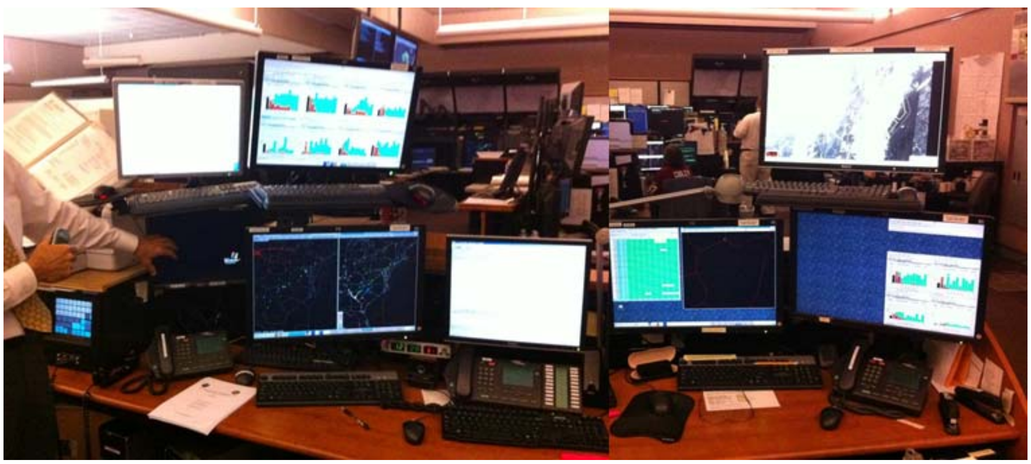


Figure 1. Supervisory Traffic Management Coordinator (STMC) position at one ARTCC TMU.

The STMC workstation shown in the photograph includes 8 displays, 7 keyboards, and 3 telephones, as well as paper and other office supplies. In addition, the workstation is located such that the STMC can see other TMU displays from the workstation. In all, 23 ARTCC STMC tools were identified, 13 of which are constantly in use. Similarly, 23 TRACON STMC tools and 19 ATCT traffic management tools were identified, including many tools that are used in multiple air traffic management domains.

While all of these tools are not deployed at every ARTCC, Borgman and Smith (2010) reported as many as 10 and as few as 3 displays at STMC workstations in other ARTCCs and discussed numerous other tools in use as well. Some of the variation in tools at a given STMC workstation is due to the tools available at a given ARTCC as well as the ability for the STMC to see displays at other workstations.

Identify Work Process(es) Using Each TMU Tool

Once the tools in the TMU were identified, the work processes they supported were identified based on observations and interviews in situ and interviews with additional subject matter experts. These work processes were characterized according to the kind of task performed. Table 1 shows the types of STMC work processes identified and the number of tools used to support the STMC in carrying out each type of work process. Two tools were counted twice. The National Traffic Management Log (NTML) supports both communication and tracking and analyzing traffic data. The Route Management Tool (RMT) acts as a database of alternate routes and also supports tracking and analyzing traffic data. Note that many of the tools used by the STMC also are used at other positions.

Table 1.
Types of ARTCC STMC work processes supported by TMU tools and the number of tools supporting each type of work process.

Purpose	Number of Tools
Communication	5
View Traffic	2
Manage Traffic	3
Weather	3
Database	3
Track and Analyze Traffic Data	9

Many work processes are supported by multiple tools. However, each has a different purpose, such as the tools for viewing traffic: the Display System Replacement (DSR) See-All Display allows TMU personnel to view the data shown on the scope of any facility controller and the Traffic Situation Display (TSD) provides a view of aircraft locations in larger areas.

Although each tool provides TMU personnel different information, they are not well integrated with each other. As a result, users must manually transfer information from one system to another. Some systems allow electronic transmission of data but do not allow automatic transmission and reporting. However, emerging technologies provide opportunities to develop data exchange standards and capabilities to support integration of the various TMU tools.

Leveraging Emerging Technologies to Improve Data Exchange

The number of tools and displays that TMU personnel must consult to gather information, make a decision, and implement an action is excessive. There is an obvious need to address the diversity of un-integrated TMU tools and displays and ensure that future capabilities relieve, rather than exacerbate, the display overload problem.

The FAA is deploying new technology such as System Wide Information Management (SWIM) as part of its air traffic control modernization program (FAA, 2012). SWIM is a data exchange interface that can be used to provide a standard that must be achieved by all TMU tools. A SWIM interface layer can facilitate data exchange among existing tools before major upgrades associated with the NextGen modernization program are available. The interface layer also can facilitate integration of additional capabilities such as traffic management course of action analysis and planning (Brinton, 2011).

The Air Traffic Course of Action Planner (ATCOAP) concept utilizes a hierarchical task network based on an operational analysis of traffic management tasks to identify and track traffic flow management issues in real-time, facilitate cross-domain coordination, and support decision-making based on integration of data from the variety of systems present in the TMU. It supports TMU personnel in identifying issues, prioritizing tasks, executing traffic management actions, and monitoring their results, as well as coordinating with other facilities in achieving these goals.

Some cross-domain coordination can be better facilitated by automating workflow and task assignment, tracking and feedback. Tasks required to achieve courses of action developed collaboratively are identified and included in a task list that is coordinated among all relevant NAS operators. The ATCOAP allows users to accept, initiate, decline or assign tasks to another operational position. Task ownership is communicated to all actors that have a 'need to know' about a given task. Figure 2 provides a graphical example of how this task tracking capability can be implemented in the ATCOAP system.

Position: ZOB TMU		
Task	**Status**	**Owner**
Re-Route Flights from ZOB37	In Progress	ZOB TMU
Implement GDP for ORD	Recommended	ATCSCC
Reduce Miles-in-Trail from ZID	In Progress	ZOB TMU
Implement Playbook LUTHE3	Completed	ATCSCC
Re-Route Airborne J32 Flights Over LUTHE	Assigned (by ATCSCC)	ZOB TMU
De-Combine ZOB22 and ZOB19	Declined	ZOB Area 1
Handle Lost Comm Aircraft in ZOB48	In Progress	ZOB Area 4
ZOB Area 2 End of Shift Log	Required	ZOB Area 2
Start Training on ZOB57	Recommended	ZOB Area 3

Figure 2. An Example of task tracking in the ATCOAP system.

In addition to developing a data exchange layer to integrate existing tools, care must be taken to ensure that emerging tools expected to support cross-domain integration also are integrated with other traffic management tools targeting the same users. Data exchange standards and plug-and-play capabilities are more appropriate for this purpose than assuming that any one tool will satisfy all TMU needs.

Conclusion

Traffic management personnel have a large number of tools to support their work. However, these tools are not well integrated with each other and therefore much TMU effort is spent in mentally integrating data from a variety of displays and manually transferring data from one system to another. A lightweight SWIM-based data exchange layer can help reduce the data entry workload of TMU personnel. It also can reduce the number of TMU displays by allowing related information to be displayed together. Such a data exchange layer also can support integrating additional capabilities to support traffic management coordination and decision making.

Acknowledgements

The authors would like to acknowledge the support of Mr. Phil Knapp, Mr. Steve Lovett and Mr. Jay Kuehne for their subject matter expertise and assistance during this effort. The authors would also like to express gratitude to NASA for funding this research effort under a Small Business Innovative Research (SBIR) contract.

References

Annett, J. (2004). Hierarchical task analysis. In D. Diaper, & N. A. Stanton (Eds.), *The Handbook of Task Analysis for Human-Computer Interaction* (pp. 67-82). Mahwah, NJ: Lawrence Erlbaum Associates.

Borgman, A. D., & Smith, P. J. (2010). *The integrated management of airport surface and airspace constraints for departures: An observational study of JFK, EWR, and IAH.* Columbus, OH: The Ohio State University Technical Report #CSEL 2010-09.

Brinton, C. (2011). *TFM COAP SBIR Phase I report.* Leesburg, VA: Mosaic ATM, Inc.

Federal Avaiation Administration. (March 2012). *NextGen implementation plan.* Washington, DC: Federal Aviation Administration.

Federal Aviation Administration. (2008). *Order JO7210.3V Facility operation and administration.* Washington, DC: Federal Aviation Administration.

Federal Aviation Administration. (2010). *Order JO7110.65T Air traffic control.* Federal Aviation Administration, Air Traffic Organization. Washington, DC: Department of Transportation.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science, 19*(3), 265-288.

Nadler, E. (2005). *Human factors integration challenges in the Traffic Flow Management (TFM) environment.* US Department of Transportation Research and Innovative Technology Administration, John A. Volpe National Transpportation Center. Washington, DC: Federal Aviation Administration Office of Aviation Research.

Nemeth, C. P., Cook, R. I., O'Connor, M., & Klock, P. A. (2004). Using cognitive artifacts to understand distributed cognition. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 34*(6), 726-735.

Woods, D. D., & Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering.* Boca Raton, FL: CRC Press.

HIGH FIDELITY SIMULATION AND AVIATION TRAINING TO IMPROVE PROBLEM SOLVING SKILLS AND COORDINATION

Andrea M. Georgiou, Paul C. Carlson, and Paul A. Craig
Middle Tennessee State University

As part of an ongoing collaborative project, a high fidelity simulation lab was developed to address teamwork deficiencies exhibited by newly minted aviation professionals. This lab, referred to as the Flight Operations Center-Unified Simulation (FOCUS) lab, was enhanced with greater fidelity, realistic and complex scenario triggers, and robust after action reviews. Student teams participate in a simulated work shift to further improve problem solving skills and coordination among senior-level undergraduate students.

In safety-critical industries such as aviation, safe and efficient operations require harmonious coordination and effective team performance across disciplines. Quite often, non-routine events that occur during flight operations require immediate and accurate responses from pilots, maintenance technicians, flight dispatchers, and air traffic controllers. Familiarity with the duties of other specializations, along with teamwork training in a simulated environment can result in more effective coordination. Currently, simulation training for such interfaces is rare. Abundant throughout the literature, the usefulness of simulation as an effective tool to improve team performance has been explored in many different safety-critical domains characterized by interdependent teams such as aviation, military, healthcare, and nuclear power industries (Andrew, Plachta, Salud, & Pugh, 2012; Burtscher, Kolbe, Wacker, & Manser, 2011; Howard, 2011; Salas, Bowers, & Rhodenizer, 1998; Waller, Gupta, & Giambatista, 2004). The ultimate goal of integrating simulation is to foster a learning environment which affords the unique opportunity to practice skills and develop strategies for effective teamwork (Arora et al., 2010; Bond et al., 2007). Errors made during a simulation should be viewed as an opportunity to learn how to improve one's level of knowledge and capabilities so as to not make the same mistakes when performing in real-world teams.

Exploring the usefulness of Simulation-Based Training (SBT) in an aviation collegiate environment, this research is a continuation of an ongoing collaborative project of a one of a kind high-fidelity simulation lab. The Flight Operations Center - Unified Simulation (FOCUS) lab is a replica of a regional airline operations center. Undergraduate students from six aerospace specializations interactively complete a simulated work shift as employees in various roles of dispatchers, pilots, ramp controllers, maintenance technicians, crew schedulers, and weather briefers. Routine and non-routine events occur during the simulation in which students must effectively communicate and coordinate with one another to reach solutions that adhere to federal regulations and standard operating procedures, all while running an airline as safely and smoothly as possible. Different from the traditional classroom training, participation in the technological innovative FOCUS lab immerses the students in a realistic airline operations center and provides the platform for increasing the knowledge, skills, and abilities (KSA's) necessary to work in a multi-team environment. Additionally, quantitative and qualitative research from the simulation lab is used to address best practices for communication, disruption management, teamwork, and group situational awareness.

FOCUS Lab Background and Concept

The conceptualization and construction for the Flight Operations Center-Unified Simulation (FOCUS) lab began in 2010 with initial funding from a NASA grant. The idea of the lab generated from industry feedback that it takes approximately 5-10 years for newly minted aviation professionals to understand the big picture of an airline and how their role and decisions impact the organization. Prior to the FOCUS lab, students were in educational silos in which they received the appropriate level of education in their specific specialization. For example, pilots trained mostly with other pilots in the program in a "silo" manner and learned all the necessary KSA's to fly an aircraft. The problem with this type of training is it fostered an atmosphere of isolation and failed to expose students as to how their role interacted with others within an organization. Understanding the importance of teamwork is vital to safe and efficient operations. In response to a need for comprehensive student training to better meet industry needs, the FOCUS project was designed to address teamwork deficiencies of senior-level undergraduate students.

What started with a round table, a few monitors, and tabletop discussions of various airline operational scenarios has evolved into a highly realistic replica of a regional airline operations center. The simulation includes complex and real-world scenarios with immediate feedback and ramifications of team member's decisions, real-time performance analysis, and in-depth debriefing procedures. Students are afforded the opportunity to work in a high-fidelity simulated operational environment and apply the knowledge learned in the classroom setting to work together and run a virtual airline. Various individual and team performance measures are utilized to analyze their problem-solving capabilities, coordination, and communication. After completion of the simulation lab, students participate in a robust debriefing procedure to review their performance and develop strategies for improvements for the next time they participate in the lab.

Teams are comprised of 10-13 senior-level undergraduate aerospace students from various specializations and they must use all available resources to effectively communication and coordinate together to meet the organizational goals of the lab. The following are the distinct positions in the lab: flight operations coordinator, flight tracking and schedule management, fuel and cargo management, weather and forecasting, crew scheduling, pseudo pilot, CRJ simulated flight crew, maintenance planning, maintenance control, and ramp tower. There are 4 sites that coordinate together during the simulation. Located on the university campus, the main operations center houses a majority of the positions and those students can interact with face-to-face communication or other methods they decide as a team. There are two more sites on the campus but in different locations which are the ramp tower for gate-to-gate management of the aircraft in the simulation and the pseudo pilot position for requesting dispatch releases prior to departures. The final simulation site is based at the local airport for the flight crew in the CRJ-200 flight training device (FTD).

Implementation of Scenario Triggers

With the overall goal to provide individuals with opportunities to practice working together as a team, scenarios must be carefully crafted in a manner that allows such opportunities. Most often, people reach out to Subject Matter Experts (SME) when developing scenarios to ensure the simulation training is an accurate reflection of the real-world environment (Georgiou, Littlepage, & Henslee, 2011). The research team for the FOCUS lab actively consults with industry experts, including line mechanics, flight dispatchers, pilots, and aerospace professors to build a library of realistic triggers for complex scenarios. Lazzara et al., (2010) explains individuals writing the scripts for scenarios need to design them in a way that promotes the KSA's identified in the performance competencies. Supporting this position, all the scenarios are designed in a manner that requires problem solving, coordination, and communication among various members within the team.

Vital for successful implementation of the scenario triggers, all teams are given scenarios that mimic real-world problems and disruptions that must be solved quickly and with minimal impact to the airline. As the FOCUS research team monitors the simulations, the triggers they issue have the potential to cause compounding downstream impacts if not handled correctly. The actions and decisions of the team determine how the research team will formulate the direction of the scenario. This allows for a great degree of flexibility and realism as students will learn that their decisions impact the degree of safety and economics of the virtual airline.

High-Fidelity Components of the Simulation

The FOCUS lab relies heavily on technology to achieve the level of realism and interactivity required by project goals. The degree to which a simulator replicates reality is referred to as fidelity and can improve the external validity of an experiment in the ability to extend the results from the simulation lab to actual working conditions (Leedy & Ormrod, 2010; Beaubien & Baker, 2004). Since a facility of this kind was unique, many technologies had to be adapted and merged to form a functional system. These technologies are comprised of both hardware and software components, some of which are commercially available, while others were customized or developed for lab-specific purposes.

All stations in the lab utilize desktop-class computer workstations with dual screen monitors. The monitors provide the needed desktop space to display multiple sources of information management and communication applications. Headsets are also available at each station, allowing for direct hands free voice communication to other workers. Above the central control desk on opposite walls lie a bank of three large LCD displays that show common-use weather maps, flight tracking radar, and flight status boards. An advanced matrix switch is in place between the source computers and the displays, allowing users to reassign screens as needed. Adjacent to the main control room is the ramp tower room. This room houses 12 computers, three large display screens, and various control stations. This equipment supports the core simulation of flights progressing along flight routes within the service area. It also depicts a photo-realistic 150-degree view of the terminal ramp environment of the hubs used by the virtual airline.

An FAA-certified professional flight training device (FTD), housed at a satellite campus training center, replicates a Canadair Regional Jet (CRJ) 200 series aircraft. As pilots fly the simulator, it electronically interfaces to simulation equipment in the FOCUS lab over a proprietary network architecture and communication protocol. As a result, the CRJ200 simulator appears on radar tracking screens and visual displays just like the other computer generated flights. The flight crew can also communicate with any station in the operations center to solve problems and share information.

Both faculty observers and students utilize a host of software applications to manage information and communicate effectively. First, Skype is used to host single and multi-party voice and text-based communication between stations. Skype is widely available, flexible, and runs on multiple platforms, so it was a clear choice for project use. Another internet application, 'join.me' is used to electronically duplicate station screens so observers can monitor individual performance on any internet enabled device.

To support their individual roles, students use Excel-based operations 'modules'. These modules were developed specifically to help students manage and interpret data for their jobs. An airline back-story is embedded in the file, that each module references, like crew duty times, flight passenger manifests, weather conditions, and cargo loads. This data can then be manipulated and shared with other stations that may need this information. The file also contains a detailed flight status board that automatically displays flight times, status, and routes. Additionally, the flight status board calculates the team's performance, indicated by average arrival and departure delays, and financial calculations.

The After Action Review (AAR) Debriefing Process

The evaluation process is of utmost importance as it facilities the feedback process. Simulations require robust debriefing protocols to provide for valuable feedback to team members (Lazzara et al., 2010; Shapiro et al., 2008). Rapid and accurate feedback is important to reinforce the lessons learned during simulated training sessions and help team members learn from their mistakes and develop action plans for future improvements (Beaubien & Baker, 2004; Hoffman, 2010; Petranek, 2000). Often with some type of debriefing procedure, team members receive feedback which should link to the learning outcomes (Shapiro et al., 2008). The Federal Aviation Administration recognizes the importance of feedback during debriefing of LOFT training (FAA, 2004). Flight crewmembers must participate in line oriented flight training (LOFT) simulations with an entire crew and all segments of flight. After completion of the videotaped simulations, the crews are debriefed and provided feedback. According to the FAA (2004), the feedback from the LOFT simulations is an excellent way for flight crewmembers to assess their skills as individuals and as team members. Based on the literature, providing feedback after simulation is the most important component that leads to improvements in team performance.

After completion of the FOCUS lab simulation, students participate in a debriefing session, referred to as the After Action Review (AAR). In this session, expert Industrial and Organizational psychologists facilitate the discussion to include identification of the positive and negative outcomes of the simulation, along with the behaviors that led to those outcomes. If there were any violations to federal aviation regulations or standard operating procedures it is discussed by the AAR facilitator to ensure these same mistakes are not made when they perform in real-world aviation teams. Students are encouraged to be professional, open, honest, and focus on the situations not identify team members with weak performance areas. Most often, during the AAR the teams develop objectives and strategies as to how to improve for the next simulation sessions. Considerations such as safety, adherence to regulations and procedures, and financial impact are the most frequent areas in which students focus on improving.

Discussion

The FOCUS lab is a powerful tool, as both a training facility and research platform. Students gain valuable experience working with others from different specializations to solve complex, time-sensitive problems. The merits of simulation to improve team performance are evident in the ability to measure and evaluate the knowledge, skills, and abilities (KSA's) of individuals interacting in a team environment (Alinier, Hunt, Gordon, & Harwood, 2006; Beaubien & Baker, 2004; Lazzara et al., 2010; Shapiro, 2008). From the data collection and analysis, feedback from simulations helps assess individual and team member skills and performance (FAA, 2004; Lazzara, 2010). Furthermore, team feedback, especially within the team, reinforces critical problem solving, teamwork, and group situational awareness. For faculty, the process of evaluating student performance often reveals potential educational deficiencies that can be utilized to influence future curriculum. The high level of realism students experience in the FOCUS lab increases participant adoption of assigned roles, thereby reducing possible disparities in motivation and 'buy-in'. High fidelity simulation benefits from interactivity, detailed user interfaces, and real-world scenarios and consequences. The After Action Reviews cement positive behaviors and learning experiences, while identifying and addressing areas of needed improvement.

Acknowledgements

This research was supported in part by a contract (NNX09AAU52G) from NASA awarded to the Middle Tennessee State University Center for Research on Aviation Training. We are grateful to Paul A. Craig at the Department of Aerospace and the Center for Research on Aviation Training at Middle Tennessee State University for his helpful advice. Correspondence concerning this article should be addressed to Andrea M. Georgiou Department of Aerospace, MTSU Box 67, Middle Tennessee State University, Murfreesboro, TN 37132. E-mail: a.georgiou@mtsu.edu

References

Alinier, G., Hunt, B., Gordon, R. & Hardwood, C. (2006). Effectiveness of intermediate-fidelity simulation training technology in undergraduate nursing education. *Journal of Advanced Nursing 54*(3), 359–369.

Andrew, B., Plachta, S., Salud, L., & Pugh C. M. (2012). Development and evaluation of a decision-based simulation for assessment of team skills. *Surgery, 152*(2), 152-157.doi:10.1016/j.surg.2012.02.018

Arora, S., Lamb, B., Undre, S., Kneebone, R., Darzi, A., & Sevdalis, N. (2010). Framework for incorporating simulation into urology training. *BJU International, 107*(5), 806-810. doi: 10.1111/j.1464-410X.2010.09563.x.

Beaubien, J. M., & Baker, D. P. (2004). The use of simulation for training teamwork skills in health care: How low can you go? *Quality and Safety in Health Care, 13*(1), i51-i56, doi: 10.1136/qshc.2004.009845.

Bond, W.F., Lammers, R.L., Spillane, L.L., Smith-Coggnins, R., Fernandez, R., Reznek, M.A., Vozenilek, J.A., & Gordon, J.A. (2007). The use of simulation in emergency medicine: A research agenda. *Academic Emergency Medicine, 14*, 353-364. doi:10.1197/j.aem.2006.11.021.

Burtscher, M.J., Kolbe, M., Wacker, J., & Manser, T. (2011). Interactions of team mental models and monitoring behaviors predict team performance in simulated anesthesia inductions. *Journal of Experimental Psychology: Applied, 17* (3), 257-269.

Federal Aviation Administration [FAA]. (2004). *Advisory circular 120-5E Crew resource management*. Retrieved from http://rgl.faa.gov/Regulatory_and_Guidance_Library/rgAdvisoryCircular.nsf/list/AC%20120-51E/$FILE/AC120-51e.pdf.

Georgiou, A., Littlepage, G., & Henslee, J. (2011). Development of criterion measures to assess interpositional knowledge and task mental models. *Proceedings of the 2011 16th International Symposium on Aviation Psychology.*

Hoffman, R.R. (2010). Some challenges for macrocognition measures. In E. Patterson & J. Miller (Eds.), *Macrocognition metrics and scenarios. Design and evaluation for real- world teams.*(pp. 11-28). Burlington, VT: Ashgate.

Howard, C. E. (2011). Simulation and training: Expecting the unexpected. *Military & Aerospace Electronics, 22*(11), 12-23.

Lazzara et al. (2010). Team Medss: A tool for designing medical simulation scenarios. *Ergonomics in Design: The Quarterly of Human Factors Applications, 18*(11), 11-17. doi:10.1518/106480410X12658371678435.

Leedy, P.D., & Ormrod, J.E. (2010). *Practical research. Planning and design* (9th ed.). Saddle River, NJ: Pearson Education.

Petranek, C.F. (2000). Written debriefing: The next vital step in learning with simulations. *Simulation Gaming, 31*(1), 108-118, doi: 10.1177/104687810003100111.

Salas, E., Bowers, C.A., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology, 8*(3), 197-208.

Shapiro, M.J., Gardner, R., Godwin, S.A., Jay, G.D., Lindquist, D.G., Salisbury, M.L., & Salas, E. (2008). Defining team performance for simulation-based training: Methodology, metrics, and opportunities for emergency medicine. *Academic Emergency Medicine, 15*, 1088-1097, doi: 10.1111/j.1553-2712.2008.00251.x.

Waller, M.J., Gupta, N., Giambatista, R.C. (2004). Effects of adaptive behaviors and shared mental models on control crew performance. *Management Science, 50*(11), 1534-1544.

TEAMWORK AND PERFORMANCE OUTCOMES OF HIGH-FIDELITY AIRLINE OPERATIONS CENTER SIMULATIONS

Glenn Littlepage, Paul Craig, Michael Hein, Richard Moffett III, Emily Sanders,
Paul Carlson, Artyom Ivakh, and Andrea Georgiou
Middle Tennessee State University

This study describes effects of a series of simulations of an airline flight operations center and related functions such as pilots and airport ramp control. It is continuation of a multi-year project designed to develop an effective training program for entry level aviation professionals and to develop insights into teamwork processes. Results indicate that the interdependent multi-specialization simulations enhanced teamwork and performance.

A variety of highly-trained professionals are needed to operate an airline. These include various specializations such as: pilot, flight dispatch, weather, crew scheduling, ramp control, and maintenance. Not only must these persons be proficient in their respective specializations, they must work together in a coordinated manner. Effective teamwork is essential for optimal airline operations and this is especially true when non-routine events occur (DeChurch & Marks, 2006; Marks, Mathieu, & Zacarro, 2001; Salas, Sims, & Burke, 2005). Previous research suggests that airline professionals in training do not have high levels of interpositional knowledge that may be needed for high levels of coordinated action (Littlepage & Henslee, 2011). Poor coordination can lead to costly flight delays and can compromise safety. Flight and delays and cancellations result in inconvenienced passengers, disruptions in business activities, and operational problems for airlines. The annual negative economic impact of delays and cancelations in the U.S. is estimated to exceed $31 billion (NEXTOR, 2010). Safety is greatly affected by a lack of coordination among aviation specializations. Merket, Bergondy, & Salas (2000) examined military accident records and determined that 68% of serious (class A) mishaps involved aircrew coordination problems. While accidents and disruptions cannot be eliminated entirely, more effective coordination among differing specializations of aviation professionals can decrease their frequency, duration, and impact.

In order to achieve these objectives, we utilize a series of high-fidelity simulations that incorporates both routine and non-routine work situations. We also incorporate after-action reviews to enhance the beneficial impacts of the simulations. Theory and research on group processes, group/team performance, and multiteam systems provides perspectives from which to view the coordination required to maintain efficient airline performance.

Effective teamwork is important for optimal levels of group performance. Multi-person simulation that captures the essential task and teamwork functions is considered to be an effective approach to team training (Howard, 2011; Salas, et al., 1998, 2010). Despite the fact that training on complex tasks facilitates transfer, most of the training literature tends to focus on relatively simple, routine tasks rather than complex, dynamic tasks requiring adaptation (Kozlowski, et al., 2001). In the current research program, we utilize individual task instruction followed by high fidelity simulations and after-action reviews in an attempt to increase awareness of interdependencies and to enhance teamwork. We do this using complex, high-fidelity simulations of the operations of a regional airline. Like actual airlines, coordination is required among various specializations. To further enhance the effectiveness of training, we utilize a series of facilitated after-action reviews. Recent research indicates that after-action reviews can lead to enhanced team performance (Villado & Arthur, 2013). The goal of this research program is to demonstrate that a training program based on these accepted practices can be used to increase awareness of interdependencies and to enhance teamwork among entry-level aviation professionals.

Method

Participants

Participants were assigned to one of 10 teams of approximately 10 persons per team. All were Aviation Science students enrolled in a senior level capstone course. The following specializations were represented: professional pilot, flight dispatch, maintenance management, aerospace administration, and aviation technology. Each of the students had completed extensive coursework in his or her respective specialization, but had relatively little knowledge of the other specializations and had limited experience working with other specializations.

Setting

The research was conducted in a four room lab facility. This facility includes a flight operations center, a ramp control tower, a pseudo-pilot room, and a CRJ flight simulator. This setting mirrors the task environments of a regional airline.

The **flight operations center** houses multiple workstations. The *flight operations coordinator* has the most central role and has final decision making authority for most matters related to the operation of flights. In order to make effective decisions, this person must utilize information from all other participants in the simulation including those in other rooms. The other positions in this room insure that flights are properly loaded and have sufficient fuel, consider weather conditions, insure that flight crews do not exceed legal duty time limitations and insure that aircraft are properly maintained and safe for flight. Personnel in the flight operations center are seated around the rim of a double row of long tables. This allows for face-to-face communication between these positions, but headset and text communication are also available. Six large screen monitors are wall-mounted, three behind each long side of the tables so that each side displays the flight schedule, a radar view of the flights in the air, and a weather map. These displays provide real-time information during the simulation. In addition, each work station is equipped with a computer and two monitors to allow access to multiple sources of data relevant to that position.

The **ramp control tower** is in an adjacent room and simulates the operation of one of the airline's hub airports (Nashville). The *ramp control specialist* directs arriving and departing flights to appropriate taxiways and gates and provides notification when a plane is ready for pushback. This room contains three wall-mounted large screen monitors providing panoramic real-time views of the gates, runways, and taxiways. Another display shows a radar view of flights preparing for landing and takeoff, and a computer allows the operator to direct the planes to gates and taxiways. The ramp control specialist can communicate with the flight operations coordinator via headphone and text.

The **pseudo-pilot room** consists of a single workstation where flights from locations other than Nashville are started. (As described above, flights from the Nashville airport are requested from the ramp tower)The *pseudopilot* can also divert aircraft if so directed and can report any issues that arise with these flights. This location consists of a single computer station. Voice and text communication are available with positions in the flight operations center and with the ramp control tower.

The **CRJ-simulator** is a fully functional simulation of a CRJ-200 cockpit area located off site. Voice and text communication is available with the flight operations center and the ramp control tower.

Procedure

The procedure consisted of a series of steps: orientation, task-specific training, initial simulation, initial after-action review, second simulation, a second after-action review, a third simulation, and a third after action review. Data were collected during four semesters. Procedures differed somewhat between the 2011 and 2012 academic years. During 2011 each team participated in two simulations and the CRJ simulator was not utilized. During 2012, the CRJ simulator was utilized and a third simulation and after-action review were added. Each of the training components is briefly described below. Following most training components, participants completed research instruments at individual computer stations. All sessions involved the operation of the same airline routes with the same resources, but weather conditions varied and events such as maintenance problems or other issues were manipulated. The simulations were designed to create the feel of a work shift much like participants will experience upon entering the workforce in commercial aviation.

Orientation. A 45-minute presentation and discussion provided a description of the lab and the various work roles. At the conclusion of the orientation, participants were informed of their team assignments and schedule of training activities.

Task Specific Training. During this 45 minute to one-hour session, each team was taken into the lab and each member was provided with instruction, demonstration, and opportunity to practice at his or her work station. The purpose of this session was to ensure that each participant developed an understanding of his or her role, responsibilities and the technical knowledge to do the job.

Simulation One. During this 2.5 hour simulation, the participants collectively worked to operate the simulated airline. The airline is a regional carrier with a fleet of 30 aircraft, two regional hubs, and 14 additional airports. During the simulation, approximately 60 flight events (takeoffs and landings) occurred. Much of the activity involved routine handling of flights and required communication and teamwork. In addition, unexpected events (such as severe weather, maintenance issues, or other problems requiring attention) occurred and further increased the need for information transfer, coordinated action, and adaptation.

After-Action Review One. Following the first simulation, participants individually completed a form about successful and unsuccessful events and reasons for these. In a follow-up session, the group participated in a facilitated discussion of positive and negative events and opportunities for improvement. This session typically lasted approximately one hour.

Simulation Two. This simulation was similar to the first simulation. It involved the same flight schedule, but a different set of problems arose.

After- Action Review Two. The second after-action review followed the same format as the first one.

Simulation Three. This simulation was similar to the first two simulations. It involved the same flight schedule, but a different set of problems arose.

After- Action Review Three. The third after-action review followed the same format as the first two.

Measures

Teamwork. Teamwork was assessed using a 30-item self-report teamwork scale developed by Mathieu and Marks and based on Marks, Mathieu, & Zaccaro, 2001. Each item was phrased as the extent to which the team actively worked to do various teamwork behaviors; the response scale ranged from 1 (not at all) to 10 (to a very great extent). The scale yields an overall teamwork score and scores for teamwork during action and transition (planning and reflection) phases and interpersonal behavior. Observer ratings of teamwork were also collected using a 10 item scale reflecting problem solving, coordination, and information utilization as well as an overall teamwork score. These measures were collected following each simulation.

Interdependence Questionnaire. This four item scale was developed to reflect facets of task interdependence (Wageman, 2001). For six positions within the simulation, participants rated the extent agreement that: their job depended on that position, that position depended on them, they had common goals, and they were part of the same team ($\alpha = .86$). Response categories ranged from 1 (strongly disagree) to 10 (strongly agree). Data were collected following the initial orientation and following the second simulation.

Communication Frequency and Communication Importance. These constructs were assessed by participant ratings of the frequency and importance of communication with seven positions within the simulation. Scales were completed following training (expected communication) and following each simulation. Frequency was measured on a 5 point scale anchored by 0 (never) to 4 (more than 10 times). Importance ratings were anchored by 0 (not at all) to 4 (absolutely essential).

Team Performance. Performance was measured by delay time during each simulation. This reflects the total hours of delays pooled across all flights scheduled during the simulation.

Results

Data were examined using repeated measures ANOVAs contrasting pre-training and post-training measures and follow-up tests as needed. See Table 1 for means and standard deviations.

Teamwork. Analysis of the self-report teamwork scale indicated that training improved overall teamwork ($p < .01$, $\eta^2 =.16$). Teamwork improved for action, $p < .001$, $\eta^2 = .10$, transition, $p < .001$, $\eta^2 = .16$, and interpersonal processes, $p < .001$, $\eta^2 = .13$. Examination of observer ratings of teamwork also revealed large improvements in teamwork. Observer ratings of teamwork increased for the overall rating, $p < .001$, $\eta^2 = .83$, problem solving, $p < .001$, $\eta^2 =.80$, coordination, $p < .001$, $\eta^2 = .85$, and information utilization, $p < .001$, $\eta^2 = .80$. These results indicate that experience working on highly interactive simulations involving specialized professional roles leads to improved teamwork.

Interdependence. Interdependence ratings were made following the first and last simulations and examined using a pre-post by position repeated measures ANOVA. This analysis did not yield a significant main effect for pre vs. post simulation, but did yield significant effects for position, $p < .001$, $\eta^2 = .55$, and the interaction, $p < .001$, $\eta^2 = .15$. Follow-up tests indicated only one change: ratings of interdependence with flight operations were higher following simulation 2 ($M = 8.36$, $SD = 1.78$) than following simulation 1 ($M = 7.87$, $SD = 2.03$), $p = .004$, $\eta^2 = .05$.

Extent of interdependence may vary across positions; for example, weather may be more interdependent with pilots than with maintenance. Since the previous analyses reflect the ratings of all participants, we conducted more refined analyses that examined interdependence between specific pairs of positions. For participants serving as flight operations coordinator, ratings of interdependence increased across simulations for the positions of: maintenance, crew scheduling, and weather. For flight operations data positions, interdependence ratings with flight operations increased while ratings with pilots decreased. For participants in the crew scheduling position, interdependence ratings with flight operations increased. These findings suggest that the simulations result in a more refined understanding of the extent of interdependence between various positions.

Communication. A pre-post by position repeated measures ANOVA compared training measures of expected communication frequency with measures following the final simulation. It revealed main effects for pre vs. post ($p < .001$, $\eta^2 = .23$), position ($p < .001$, $\eta^2 = .75$), and the interaction ($p < .001$, $\eta^2 = .42$). Additional analyses revealed that communication frequency decreased for five positions, but did not change for the flight operations coordinator or maintenance positions. Similar analyses were conducted for communication importance. Significant effects were observed for pre vs. post ($p < .001$, $\eta^2 = .45$), position ($p < .001$, $\eta^2 = .84$), and the interaction ($p < .001$, $\eta^2 = .65$). Additional analyses indicated that ratings of communication importance decreased for five positions, but remained stable for the flight operations coordinator position. These findings that the simulations result in an overall decrease in communication and in communication patterns that are more focused toward the key position of flight operations coordinator.

Delay Time. Delay time decreased in 15 of the 17 teams (88%). Mean delay time decreased by more than five hours, ($p = .015$, $\eta^2 = .32$). These results indicate that team performance improved as training progressed.

Table 1

Measure	**Initial**	**Final**
	Mean (*SD*)	Mean (*SD*)
Overall (Self-Rated) Teamwork	3.68 (.70)	3.97 (.64)
Action Processes	3.59 (.76)	3.85 (.72)
Transition Processes	3.50 (.85)	3.88 (.75)
Interpersonal Processes	3.97 (.73)	4.23 (.65)
Overall (Observer-Rated) Teamwork	3.03 (.45)	4.37 (.64)
Problem Solving	2.88 (.44)	4.25 (.70)
Coordination	2.97 (.47)	4.38 (.65)
Information Utilization	3.24 (.48)	4.49 (.61)
Interdependence	7.64 (1.67)	7.72 (1.55)
Communication Frequency	2.42 (.74)	1.65 (.74)
Communication Importance	2.95 (.62)	2.16 (.88)
Delay Time (hours)	11.33 (5.11)	6.31 (5.47)

Discussion

The reduction in delay times indicates that team performance improved as a result of the training. It is likely that some of the performance gain resulted from improved individual knowledge and skill, but our other findings suggest that another reason why performance improved is that participation in the simulations and after-action reviews allowed participants to learn to work more effectively as a team. These findings suggest that high-fidelity simulations of complex tasks can lead to enhanced teamwork, an increased and more intricate awareness of interdependencies, a reduction in overt communication, and a more differentiated pattern of communication to specific positions. Additional work is planned to examine the effects of training on emergent cognitive states and to examine the relations of processes and cognitive states to team performance. The current research extends work on team training by showing that high-fidelity simulations can facilitate teamwork and awareness of interdependence in cross-functional teams conducting extremely complex tasks

References

DeChurch, L. A., & Marks, M. A. (2006). Leadership in multiteam systems. *Journal of Applied Psychology, 91*, 311-329. doi:10.1037/0021-9010.91.2.311

Howard, C. E. (2011). Simulation and training: Expecting the unexpected. *Military and Aerospace Electronics, 22*(11), 12-23.

Kozlowski, S. W. J., Toney, R. J., Mullins, M. E., Weisband, D. A., Brown, K. G., & Bell, B. S. (2001). Developing adaptability: A theory for the design of integrated-imbedded training systems. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research* (Vol. 1, pp. 59-123). Amsterdam: JAI/Elsevier Science.

Littlepage, G. E., & Henslee, J. A. (2011). Multiteam coordination in simulated airline operations: Assessment of interpositional knowledge and task mental models. *Proceedings of the 16th annual symposium on aviation psychology*, 603-608.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review, 26*, 356-376. doi:10.2307/259182607.

Merket, D.C., Bergondy, M. L., & Salas, E. (2000). Making sense out of team performance errors in military aviation environments. *Transportation Human Factors, 1*, 231-242.

Nextor. (2010). Total delay impact study: A comprehensive assessment of the costs and impacts of flight delay in the United States. Retrieved February 3, 2011 from: http://www.nextor.org/pubs/TDI_Report_Final_11_03_10.pdf

Salas, E., Bowers, C.A., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology, 8*(3), 197-208.

Salas, E., Cooke, N. J., & Gorman, J.C. (2010). The science of team performance: Progress and the need for more…. *Human Factors, 52*, 344-346.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "big five" in teamwork? *Small Group Research, 36*, 555-599. doi:10.1177/1046496405277134.

Villado, A. J., & Arthur, W., Jr. (2013). The comparative effect of subjective and objective after-action reviews on team performance on a complex task. *Journal of Applied Psychology*. Advance online publication. doi: 10.1037/a0031510.

Wageman, R. (2001). The meaning of interdependence. In M. E. turner (Ed.). *Groups at work: Theory and research*, pp. 197-217. Mahwah, NJ: Erlbaum.

REFINEMENT OF MENTAL MODELS OF INTERDEPENDENCE AND COMMUNICATION

Emily Sanders, Glenn Littlepage, and Michael Hein
Middle Tennessee State University

This study examined the relationship between mental models of interdependence and communication in a simulated aviation flight operations center. Social network analysis indicated that mental models of both interdependence and communication importance develop with team interaction and that, following team interaction, the two types of models were closely related.

Increasing complexities involved in the world of work have forced companies to turn to teams in order to meet the new challenges (Smith-Jentsch, Campbell, Milanovich, and Reynolds, 2001). In the world of aviation, flight operation centers incorporate numerous team members to integrate complex information and make decisions in real-time. Various positions are needed in this work group, and each has unique information that needs to be communicated for effective decision-making and problem solving.

Flight operations centers are the hub of communication, information, and coordination among various critical aviation specialties. It is here that the converging disciplines must work together to ensure the safe and efficient operations of each individual aircraft and the overall airline. With the magnitude of information needed to accurately and efficiently make decisions in this industry and with the high consequences of errors, airlines need teams to meet the growing demands of information processing, technology, and high-stakes decision-making. Researchers have argued that these teams should be better equipped to function in this industry by sharing the burden of the tasks, backing each other up, contributing specialized skills, increasing the available amount of knowledge, and managing themselves (Mathieu, et al., 2000; Mesmer-Mangus & DeChurch, 2009). By utilizing groups of experts, teams can filter the necessary information through the group. The group can then integrate the information from various disciplines and make the best possible decision.

Mental models

Rouse and Morris (1986) defined mental models as "mechanisms whereby humans generate description of system purpose and form, explanations of systems functioning and observed system states, and predictions of future system states." (p. 7). This definition describes the three purposes of mental models; description, explanation, and prediction. As individuals increase their interaction with the team, they begin to understand how they fit into the overall teamwork process. This enables them to explain with whom they need to communicate and what they should be doing in any given situation. This understanding will also help them predict what their team members will do, which further helps them to determine their role in each novel situation. Over time, teams develop a deeper understanding of the systems they utilize, the information needed for optimal operation, and the team members that should be contacted for specific knowledge. They understand how they must coordinate their activities and who needs to be involved in what decisions.

Mental models were first applied to the team-level by Cannon-Bowers, et al., (1993) during observations of expert systems in which it was noted that some teams outperformed others by having highly coordinated actions and behaviors, while not necessarily increasing overt communication. Recent research illustrated that a cognitive base to teamwork exists and it was been shown to impact behavior, motivation, and performance (Cooke, Gorman, Duran, & Taylor, 2007; DeChurch & Mesmer-Mangus, 2010). Mental models are essentially mental patterns and structures that enable individuals to interact with their environment through cues that enable them to predict and explain events occurring around them (Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). In flight operation centers, communication and coordination are severely impacted by time pressure, workload, and the consequences of error are great. These factors make mental models crucial, because normal communication channels are hampered by environmental constraints (Mathieu, et al., 2000; Stout, Cannon-Bowers, & Salas, 1996).

Team communication

Communication was identified as one of the key behavioral processes in teams (Kozlowski & Bell, 2003; Salas, Sims, & Burke, 2005). Communication allows teams to be able to fully tap into their collective expertise.

Communication seems to contribute to the development of shared mental models and once these models have been developed, they serve to refine and focus communications. Before a team can coordinate, they must communicate. Without communication, coordination is unable to occur (Keyton, Ford, & Smith, 2012). As team processes, such as communication, become routinized over time, they produce shared cognition. The pre-cursor to coordinated communication is a shared understanding; i.e. a shared mental model. In the case of communication, this shared mental model allows for more coordinated communication and understanding of whom to communicate with in each new situation.

Teamwork mental models of communication importance are mental understandings of the importance of various communication channels. Previous team research has indicated that with interdependent team interaction, overt communication decreases and that the decreases are specific to certain positions within the team. (Littlepage, Craig, Hein, Moffett, Georgiou, & Carlson, 2012). This suggests that communication becomes more selective and focused as a result of the development of mental models. Other research has found that high performing teams did not necessarily increase the amount of communication occurring within the team (Cannon-Bowers et al., 1993).

Social network analysis

Most research in psychology is focused on the attributes and characteristics of specified subjects. Social network analysis has a primary focus on the relational ties between actors; the actors in a study being organizations, groups, or individuals in a previously determined context (Scott, 1991). Social network analysis (SNA) maps relational ties between individual nodes, in this case team members, to produce an illustration of the entire network. SNA is especially suited to study social interaction, which is only understood in its fullest by review of the entire network. Social network analysis can be used to determine similarity of mental models among team members.

Research questions

As individuals increase their interaction with the team, they begin to understand how they fit into the overall teamwork process. This enables them to explain with whom they need to communicate and what they should be doing in any given situation. This understanding will also help them predict what their team members will do, which further helps them to determine their role in each novel situation. Participants may come into the simulation with previously set mental models due to understanding and a general knowledge of their specialization, but because team interaction influences mental model development and because mental models help individual members determine with whom they will communicate, the following research questions will be evaluated in this study.

1. Will mental models of interdependence change over time as a result of team interaction?
2. Will mental models of communication importance change over time as a result of team interaction?
3. Are mental models of interdependence related to mental models of communication importance?
4. Does any change in communication importance occur in parallel with interdependence?

Method

Sample and procedures

To test our hypotheses we evaluated multiple teams of upper-level undergraduate students placed in a simulated flight operations command center. Participation in the research was a course requirement; however, students were not mandated to allow responses to be used for data analysis. Multiple specialties were represented in each group. These concentrations consisted of professional pilot, administration, flight dispatch, technology, and maintenance. Teams varied in size, but on average there were 10 participants in each team. Each participant was assigned a position in the lab. These positions were as follows: flight operations coordinator, flight operations data weight and balance, flight operations data planning and scheduling, weather and forecasting, crew scheduling, maintenance, pilot, and ramp tower. For purposes of analysis flight operations data and flight operations coordinator were aggregated into one flight operations position.

On day one, students were given information, in on-boarding fashion, regarding the virtual airline where they were “employed” and asked to sign consent forms. During this time explanations of all positions and information regarding the layout and function of the various workstations was provided. Following the on-boarding session, students were split into work teams. Each student was assigned a position and given information specific to that position. Students then completed pre-simulation measures, including a measure of interdependence. This measure captured the participant's perception of the extent of interdependence; each participant rated the extent of agreement that his or her job depended upon each other position. Following this the students were given individual training regarding their position, a rundown of where everything was located in the lab, and the functions of each position. Students then completed another set of pre-simulation measures, including expected communication importance. This measure asked participants to indicate how important it was for their position to communicate with each of the other positions within the simulation.

After training, students participated in the high-fidelity simulations. Over the course of the semester, students participated in two or three 2.5 hour simulations. Next, participants were given the post-simulation measure of communication importance. In addition, teams participated in after-action reviews following each simulation to discuss the outcomes of the simulation, behaviors that contributed to those outcomes, and possible changes for upcoming simulation(s). After the final after-action review, students completed the post-simulation measure of interdependence.

Variables and measures

Mental models of interdependence. Mental models of interdependence are mental understanding of members' dependency on each of the other positions. The interdependence measure consisted of a question of how much the participant's job was dependent on each of the other positions. Responses were made using a 10-point Likert scale, (1 = strongly disagree, 10 = strongly agree). This question provided the basis for creating ego-net based matrices that could then be analyzed using the QAP correlation to measure the correlation with other matrices (Hanneman & Riddle, 2005). To create these matrices responses were aggregated into 6x6 or 5x5 matrices. (Ideally, all matrices would be 6x6 to account for all the functions in the simulation, but for some teams missing data resulted in 5 x 5 matrices). The QAP correlation allows researchers to analyze how two matrices converge onto each other and provides a correlation between the two networks. A measure of density was computed to indicate the extent to which all positions were perceived as interdependent.

Mental models of communication importance. The communication patterns measure asked the participants to indicate how important the communication is between their position and the various other positions. This measure is scaled on a 5-point Likert scale, (0 = Not important at all, 4 = Absolutely essential). It is taken at two separate times; prior to simulations and again after all the simulations. This question is used to build an ego net based matrix and QAP correlations were computed. For measures of density, the communication importance matrices were dichotomously coded to represent the lack of communication (0), or the presence of communication (1). Values 3 and above were coded with a 1.

Analysis

Because of missing data, the number of matrices (teams) varied across analyses from 9 to 13 teams. Correlations among mental models were assessed using UCInet (Borgatti, Everett, & Freeman, 2002). This software allows for a network of relational ties to be created, which can then be compared for a number of measures including correlation within and across networks. In this study, the networks (matrices) were created for both interdependence and communication importance at two time periods. In order to address the research questions, QAP correlations were obtained using the matrices created from the two measures. QAP analysis provides a correlation of how similar two square matrices are to each other. Significant QAP correlations indicate similarity between the two matrices and, thus, the presence of a correlation between the mental models. In addition, density was used to investigate the connectedness of communication importance and interdependence networks. For binary data, decreases in density represent a decrease in the number of ties between dyads. These ties do not take into account the direction of the relationship. For valued data, density can be thought of as the strength of the ties present between dyads (Hanneman & Riddle, 2005).

To investigate whether mental models of interdependence changed over time as a result of team interaction pre-simulation interdependence matrices were correlated with post-simulation interdependence matrices (QAP 1). The size of the QAP correlation indicates stability. Low correlations indicate change and are consistent with the development of mental models. In the same fashion, pre-simulation communication importance matrices were correlated with post-simulation communication importance matrices, QAP 2, to evaluate the changes in mental models of communication importance. Again, a lack of correlation will suggest that changes in mental models occurred. In addition to these QAPs, density was measured for both communication importance matrices (pre-simulation and post-simulation) using a dichotomized matrix, Likewise, density was measured for matrices of interdependence both pre and post simulation. Matrices of interdependence were not dichotomized.

Analyzing the correlation between pre-simulation measures for both interdependence and communication importance illustrated how closely related the two mental models were initially (QAP 3). Analyzing post-simulation measures for interdependence and communication importance displayed the relationship between the two models after the simulations (QAP 4). QAP correlations between communication measures and interdependence measures at each time period indicate the degree of similarity between mental models of interdependence and communication importance prior to and following simulations.

Results

QAP analyses 1 and 2 examined the stability of mental models from initial to post-simulation. Some teams showed significant correlations, while others did not. The mean correlation for QAP1 was .24 while the mean correlation for QAP 2 was .43. These correlations indicate that the initial mental models account for less than 19% of the variance in post-simulation mental models. This suggests that mental models may be revised as a result of team interaction.

Average pre-simulation interdependence density (7.76) was found to be higher than post-simulation density (7.34). In the same fashion pre-simulation density of communication importance (.52) was found to be higher than post-simulation communication importance (.29). A paired comparisons t-test performed on both interdependence and communication importance found that these decreases were statistically significant ($p < .05$, in both cases). This may indicate that the mental models are becoming more refined, and the networks less tangled. Although density decreased for both interdependence and communication importance, there was no significant correlation between the differences of pre and post simulation interdependence and communication importance.

Correlation between pre-simulation interdependence and communication importance (QAP analysis 3) was variable, but provided an average correlation of $r = .17$. This suggests that a weak and/ or inconsistent relationship exists between participants' understanding of communication importance and interdependence prior to simulation and interactions. QAP 4, regarding the correlation between post-simulation interdependence and communication, held mixed results for individual teams. However, it is important to note that more teams showed a significantly positive correlation at post-simulation than at pre-simulation. In addition, the average correlation increased to $r = .46$. This suggests that, following team interactions in the simulations, mental models of communication importance were more closely related to mental models of interdependence. Participants thought communication was more important between team members who were highly interdependent. The pattern of correlations between post-simulation interdependence and post-simulation communication importance warrants further research. (QAP results for each team are presented in the Appendix.)

Conclusions

In conclusion, analysis indicated that mental models may be refined as teams interact, and that little variance in post-mental models can be explained by pre-simulation mental models. In addition, density analysis illustrated a significant decrease in both models. This indicates that individuals may begin to become more aware of the positions with which they are truly interdependent, and which positions are important to communicate with. They begin to understand how their position fits into the entire network. QAP analysis 3 & 4 indicated that post-simulation, participants' mental models of interdependence and communication were more closely related than prior to simulations. This finding suggests that as participants realized how interdependent their jobs were they also realized that communication with these positions was important. Again, this allows communication to become more focused and narrowed. Taken in the context of other research (Littlepage et al., 2012), as teams continue to interact their mental models of interdependence change which influences who they perceive it is important to communicate with. Interestingly, teams perform better as simulations continue, but frequency of communication decreases. As team members begin to understand who they need to interact with in order to perform their tasks, their communication becomes more selective and coordinated, which can lead to higher performance. Taken as a whole these results imply that mental models are refined through interaction, and that as these refinements take place smaller and more unified networks are formed within the team. Although mental models of both interdependence and communication importance develop and are related, these mental models are not refined in parallel and the time of this refinement is still unknown.

Limitations and future research

This study, like all, has its limitations. Sample size (9 – 13 teams) was a limitation. Because this is a simulation, generalization is an issue. However, the situation was realistic and professionally relevant to participants. Although additional studies using larger samples of teams are needed, these results suggest that mental models of interdependence and communication importance can become more refined and more accurate as a result of team experience.

REFERENCES

Borgatti, S.P., Everett, M.G. & Freeman, L.C. (2002) Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. r. Castellan (Ed.), *Individual and group decision making: Current issues* (pp. 221-246). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Cooke, N.J., Gorman, J.C., Duran, J.L., & Taylor, A.R. (2007). Team cognition in experienced command-and-control teams. *Journal of Experimental Psychology 13*(3), 146-157.

DeChurch, L.A. & Mesmer-Mangus, J.R. (2010). Measuring shared team mental models: A meta-analysis. *Group Dynamics: Theory, Research, and Practice 14*(1), 1-14.

DeChurch, L.A. & Mesmer-Mangus, J.R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology 95*(1), 32-53.

Hanneman, R.A. & Riddle, M. (2005). Introduction to Social Network Methods. Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/)

Keyton, J., Ford, D.J, & Smith, F.L. (2012). Communication, Collaboration, and Identification as Facilitators and Constraints of Multiteam Systems. In S. Zaccaro, M. Marks & L. DeChurch (Eds.) *Multiteam Systems* (173-190). New York, NY: Taylor & Francis Group.

Kozlowski, S.W. & Bell, B.S. (2003). Workgroups and Teams in Organizations. In W. Borman, D. Ilgen & R. Klimoski (Eds.) *Handbook of Psychology: Vol. 12 Industrial and Organizational Psychology* (333-375) London: Wiley.

Littlepage, G., Craig, P., Hein, M., Moffett, R., Georgiou, A., & Carlson, P. "Training to Enhance Mutliteam Coordination in the Airline Industry" Society of Industrial Organizational Psychology [Conference] San Diego. 26. April. 2012.

Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Salas, E. & Cannon-Bowers, J.A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology 85*(2), 273-283.

Mesmer-Mangus, J.R. & DeChurch, L.A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology 94*(2), 535-546.

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*, 359 – 363.

Salas, E., Sims, D.E., & Burke, C.S. (2005). Is there a "big five" in teamwork? *Small Groups Research, 36*, 555 – 599.

Scott, J. (1991). *Social Network Analysis: A Handbook.* Newbury Park, CA: Sage.

Smith-Jentsch, K.A., Campbell, G.E., Milanovich, D.M., Reynolds, A.M. (2001). Measuring teamwork mental models to support training needs assessment, development, and evaluation: two empirical studies. *Journal of Organizational Behavior, 22,* 179-194.

Stout, R. J., Cannon-Bowers, J. A., & Salas, E. (1996). The role of shared mental models in developing team situational awareness: Implications for training. Training Research Journal, 2, 86-116

Appendix

QAP Analysis1-4.

QAP Analysis 1 (Interdependence)		
TEAM	*R*	*P*
Team 1,6	0.616	0.021
Team 3,6	0.170	0.250
Team 4,6	0.065	0.415
Team 1,7	0.663	0.000
Team 3,7	0.717	0.000
Team 5,7	-0.621	0.036
Team 6,7	0.621	0.000
Team 1,8	-0.266	0.143
Team 6,8	0.164	0.333

QAP Analysis 2 (Communication)		
TEAM	*r*	*p*
Team 1,6	0.336	0.128
Team 2,6	0.189	0.17
Team 4,6	0.344	0.081
Team 1,7	0.619	0.023
Team 2,7	0.559	0.041
Team 3,7	0.450	0.043
Team 6,7	0.606	0.018
Team 3,8	0.451	0.101
Team 4,8	0.342	0.087
Team 6,8	0.443	0.061

QAP Analysis 3 (Pre-Simulation)		
TEAM	*r*	*P*
Team 1,6	0.364	0.059
Team 2,6	0.366	0.045
Team 3,6	-0.149	0.267
Team 4,6	-0.201	0.174
Team 1,7	0.124	0.236
Team 3,7	0.339	0.082
Team 4,7	0.543	0.110
Team 5,7	0.118	0.342
Team 6,7	0.185	0.241
Team 1,8	-0.044	0.410
Team 3,8	0.001	0.486
Team 4,8	0.141	0.417
Team 6,8	0.464	0.016

QAP Analysis 4 (Post-Simulation)		
TEAM	*r*	*p*
Team 1,6	0.502	0.032
Team 2,6	0.659	0.022
Team 4,6	0.627	0.006
Team 1,7	0.389	0.159
Team 5,7	0.320	0.180
Team 6,7	0.505	0.064
Team 2,8	0.849	0.018
Team 3,8	0.267	0.178
Team 4,8	0.007	0.506
Team 6,8	0.446	0.022

A SITUATION AWARENESS DESIGN APPROACH TO THE POSITION OF AIRLINE MAINTENANCE CONTROL IN A SIMULATED OPERATIONS CONTROL CENTER

Durant Camron Bridges III
Middle Tennessee State University
Murfreesboro, TN

Research is being conducted at mid-sized public university regarding the collaborative team efforts of aviation students in a simulated airline control center. The students employ coordinated problem solving efforts toward disruption management and schedule optimization. Although the focus of this research is communication of the decision support systems, the design of the Maintenance Control position is discussed at length. The Maintenance Controller(s) communicate with all positions, as well as departments located outside the center's physical location that may and may not be accessed by other participants. This position is one of the most vital and engaging in the simulation, and its details help illustrate the myriad of research areas and opportunities that a center of this magnitude represents.

The university performing the research has recently established a training center project, which places students in a simulated airline flight operations control center. In the center, students are arranged in teams and represent the roles of the departments running an actual airline operations control center. Students from the five aviation specializations interactively complete a simulated work shift fulfilling the responsibilities of dispatchers, pilots, ramp controllers, crew schedulers, weather briefers and aircraft maintenance controllers. The teams are given realistic scenarios throughout their shifts that require the students to work together to resolve issues quickly and effectively. Each team works to meet organizational goals by focusing on safety, on-time performance, customer satisfaction, and disruption management. In order to achieve this efficiently, coordination across disciplines is required. The design of the aircraft maintenance department was designed around the role of the Maintenance Controller and the components that may expose their level of situation awareness. This design is discussed pertaining to the three levels of perception, comprehension and projection (Endsley, 1988).

Situation Awareness

"Situation awareness (SA) is the detection 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 1988). Team situation awareness is defined as "the degree to which every team member possesses the situation awareness require for his or her responsibilities" (Endsley 1989). Figure 1 illustrates the overlap of shared information from multiple team members and furthers the importance of good team SA as more than one member may require correct information in order to perform their functions within the team (Endsley, 1989).

In an article in *The International Journal of Industrial Ergonomics* about aircraft maintenance team situation awareness, Mica Endsley and Michelle Robertson (2000) describe the parameters used to measure and design a training program geared toward increasing the situational awareness of aircraft maintenance technicians and their counterparts. Endsley and Robertson began by pointing out four (4) key issues in the domain of human error that can be identified in aircraft maintenance:

(1) Shortcomings in the detection of critical cues regarding the state of aircraft or subsystems
(2) After perception of symptoms, difficulties interpreting meanings or significance regarding associated information
(3) Shared tasks of multiple individuals on one aircraft
(4) Coordination of information across shifts and between maintenance departments

The university conducting the research attends, in part, to the fourth issue of departmental cross-coordination and this paper utilizes previous research on situation awareness to describe the design on the components therein (Endsley & Robertson, 2000). The design of the aircraft maintenance department in the simulation follows the team situation awareness context analysis, which is divided into two areas: SA Requirements Analysis and SA Resource Analysis (Figure 2). The SA Requirement Analysis is comprised of three parts: SA requirements, decisions, and goals. SA Resource Analysis is composed of personnel and technology components that combine to form SA Resources.

Situation Awareness Resources: Personnel and Technology

The simulated airline maintenance department consists of two aircraft maintenance hubs with tooling, equipment and personnel (Figure 3) available to perform scheduled maintenance tasks, as well as help respond to unscheduled maintenance problems that may occur in the fleet of 30 aircraft operating over one hundred flights during work shifts. These resources are at the disposal of the student playing the role of the aircraft Maintenance Controller, who is responsible for the short-term planning of line maintenance activities during aircraft downtime between flight legs, also known as turn-around time. Additionally, they respond to operational disruptions and make GO/NOGO decisions based on aircraft physical condition assessments and additional information concerning financial and operational limitations of the fleet. Depending on maintenance task eligibility, students may elect to handle issues at the time of discovery or defer them for future addressing.

In order to perform his or her duties, the Maintenance Controller must communicate directly with each department to disseminate which information is pertinent to consider for decisions in the aircraft maintenance department. Figure 4 illustrates all possible communication paths to and from the Maintenance Control position. Several positions are located offsite and include a Pseudo Pilot that operates the majority of scheduled flight as part of a flight tracking software package, a pair of pilots operating a level 5 CRJ simulator, and representatives for the airline and airports the two aircraft maintenance hubs.

The Maintenance Control position, like all positions in the simulation, follows a paperless design that utilizes virtual applications. The primary technologies used for passing information include a commercial information system for logging maintenance activities and a commercial direct-connection application for interdepartmental VOIP calls and instant messaging. Additionally, all aircraft maintenance manuals and the minimum equipment list are provided in pdf format. Additional departments utilize a multitude of technical software packages that monitor aircraft flight tracking, weather, crew resources, and airline flights schedules.

Situation Awareness Requirements: Goals and Decisions

Specific goals of the Maintenance Control position are shown in Table 1. They are based upon those of the aircraft maintenance technicians in the team situation awareness study but limited by the role in the simulation from performing physical tasks on the aircraft (Endsley & Robertson, 2000). Table 2 illustrates the situation awareness requirements from the perspective of the role of the Maintenance Controller. This includes the goals, subgoals, major decisions and SA requirements. As stated earlier, these benchmarks may be used to determine individual SA levels, as well as the position's ability to contribute to team SA. Here, it becomes more evident the position's reliability upon relayed information pertaining to aircraft status in terms of troubleshooting and system functionality. They are also communicated a variety of information that would otherwise be automatically available in the case of a functioning part 121 operator.

Problems and Limitations

Unlike aircraft maintenance courses, which allow the student to familiarize themselves with aircraft systems and troubleshooting techniques through multiple channels, this simulation does not rely on the Maintenance Control position to personally resolve maintenance issues on the aircraft. Instead, the student is asked to make management decisions of whether or not to execute maintenance actions such as troubleshooting, deferment and scheduling/rescheduling maintenance tasks. Although some knowledge of aircraft systems is required, the subject matter involved in the simulation does not go beyond the scope of all aircraft maintenance participants, who have completed a variety of aviation maintenance courses and are at the ends of their academic programs.

References

Endsley, M.R. (1988). Design and evaluation for situational awareness enhancement. In *Proceedings of the Human Factors Society 32nd Annual Meeting,* Santa Monica, CA. 97-101.

Endsley, M.R. (1989). Final report: situation awareness in an advanced strategic mission. NOR DOC 89-32, Northrop Corporation, Hawthorne, CA.

Endsley, M.R. (1999). Situation awareness and human error: Designing to support human performance. In *Proceedings of the High Consequence Systems Surety Conference*, Albuquerque, NM.

Endsley, M.R. & Robertson, M.M. (2000). Situation awareness in aircraft maintenance teams. *International Journal of Industrial Ergonomics, 26,* 301-325.

Federal Aviation Administration [FAA]. Airworthy or unairworthy? Compliance and enforcement. Introduction to investigation and compliance-related tasks. 14-87(1). Retrieved on February 20, 2013 from http://fsims.faa.gov/WDocs/8900.1/V14%20Compliance%20&%20Enforcement/Chapter%2001/14_001_005.htm

Tables and Figures

Table 1.
Maintenance Control goals (Based upon Endsley & Robertson 2000).

- 1.0 Aircraft safety
  - 1.1 Deliver aircraft in airworthy, safe condition
    - 1.1.1 Find potential problems
    - 1.1.2 Solve problems
    - 1.1.3 Make repairs
      - 1.1.3.1 Determine part availability
      - 1.1.3.2 Placard problem
    - 1.1.4 Service aircraft
    - 1.1.5 Provide quality workmanship
- 2.0 Deliver aircraft on time
  - Prioritize tasks

Table 2.
Maintenance Control SA requirements (Based upon Endsley & Robertson 2000).

- 1.0 Aircraft safety
  - 1.1 Deliver aircraft in airworthy, safe condition
    - 1.1.1 Assess reported potential problems
      - *Item within or beyond serviceable limits?*
      - *Item near limits needing preventive maintenance?*
    - 1.1.2 Solve problems
      - *Fix problem or defer?*
        - potential impact of problem on flight safety
        - time required to solve problem
          - time required to get part
        - length of time item can be deferred without repair (MEL category)
        - location(s) aircraft is going to
          - facility maintenance capabilities
        - today's load
        - problem deferability category (placardable, groundable)
          - minimum equipment list (MEL) status
      - *Problem requires extreme action?*
        - replace aircraft
        - cancel flight
    - 1.1.3 Make repairs
      - 1.1.3.1 Determine part availability
        - *How long to get part here?*
      - 1.1.3.2 Placard problem
        - *Can problem be placarded?*
          - type of problem
          - Minimum Equipment List (MEL) status
          - Deferred information placard (MEL number)
          - Open item list (OIL)
          - redundant systems available
          - flight number

1.1.4 Service aircraft

- *Service activities needed?*
    - Tasks to be done
        - scheduled maintenance (A checks/parts replacements)
    - Current status of job?
        - status of other tasks impacting own task
        - other tasks own task will impact
        - major problems encountered

1.1.5 Provide quality workmanship

- *Activities reported performed?*
- tasks performed
- paperwork completed

2.0 Deliver aircraft on time

2.1 Prioritize tasks

- *Best order for tasks?*
    - task time requirements
    - interdependence/sequencing requirements of tasks
    - problem deferability category (placardable/groundable)
        - Minimum equipment list (MEL) status
    - availability of parts
    - availability of personnel
    - availability of tools and equipment

Figure 1.
Team situation awareness (from Endsley, 1989).

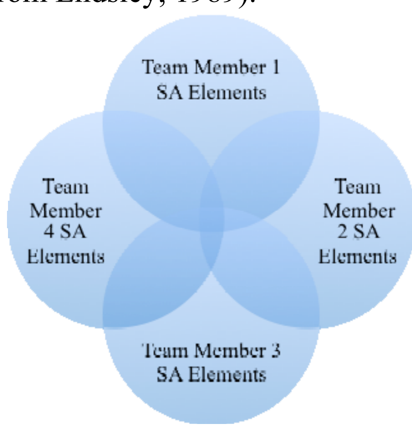


Figure 2.
Hub Maintenance Base Personnel (Based upon Endsley, Robertson 2000)

Figure 3.
Hub Maintenance Base Personnel

Figure 4.
Maintenance Control personnel SA resources.

EXPLORING THE EFFECTS OF WORKING MEMORY CAPACITY, ATTENTION, AND EXPERTISE ON SITUATION AWARENESS IN A FLIGHT SIMULATION ENVIRONMENT

Orçun Orkan Özcan, Murat Perit Çakır, and Bilge Say
Department of Cognitive Science, Middle East Technical University, Ankara, Turkey

Simulator pilots are subject to some of the constraints of a real flight situation in a PC based flight simulation. Situation awareness (SA) for simulator pilots was explored in terms of underlying cognitive aspects by analyzing the compound effects of expertise, working memory, inhibition and divided attention. Online and Offline SA measurements were analyzed with expertise and scores of Automated Operation Span Task, Stroop and Coşkunöz visual attention tasks. Regression analyses revealed the expected relationships of simulator pilots' SA with expertise and inhibition capacity but not with working memory and divided attention capacity. Obtained results were also compared to those of professional pilots. Despite similar cognitive capacities and expertise, simulator pilots had incompatible results with professional pilots in offline SA queries and they exhibited different SA performance related to expertise and cognitive capacity tests. This situation probably resulted from unsystematic differences in simulator pilots' practices.

The construct of Situation Awareness (SA) is a quite rigorous term standing for composition of many psychological abilities. Considering the cognitive aspects of the phenomenon, it can be seen that SA has been associated with several abilities like perception, long-term memory, working memory, attention, reasoning, and decision-making (Horswill and McKenna, 2004; Breton and Rousseau, 2001; Endsley, 1995b, 1997). However, the current status in the literature on SA does not converge to a well-defined combination of these abilities. This situation motivates a need for cognitive elaboration in order to clarify SA's components as a complex cognitive phenomenon. The cognitive components of SA were described by a limited number of studies (Durso & Gronlund, 1999; Endsley, 1995b; Sarter & Woods, 1991; Wickens, 1999) and individual differences in SA memory requirements were studied by few researchers (Carretta, Perry & Ree, 1996; Johannsdottir, 2004; Sohn & Doane, 2004). However, in most cases, the experimental designs of these studies either lack an explicit assessment of SA in the task environment or do not involve operators/ pilots as participants. Consequently, further explorations are needed for the cognitive grounding of SA. A recent study on this area has been conducted by Serkan Çak (2011) where 36 professional pilots were subjected to SA and cognitive capacity tests. In the current study, the flight scenario, SA queries and cognitive capacity tests from Çak (2011) were used for simulator pilots with modifications (Özcan, 2012). The research motivation was to find out the possible differences in the cognitive contributions to SA between professional and simulator pilots, who were not pilots but have been trained and experienced in PC based flight simulator environment.

Method

The experimental setup used in the study consisted of a simulated flight task, AOSPAN, Stroop and Coşkunöz visual attention tests for measurement of SA, working memory capacity, and inhibition and divided attention capacities, respectively. In order to have comparable results between professional and simulator pilots, this study was based on Çak's doctoral studies (2011) and similar experiments, except for divided attention, with modifications were used. The original flight scenario has been modified for simulator pilots together with expert simulator pilots who worked as Subject Matter Experts (SME). Multimodal divided attention task was changed with a only visual task because simulator pilots do not use radio extensively and they don't pay considerable attention to auditory modality. For this reason, Coşkunöz visual attention test was introduced. Detailed information and modifications on the original scenario can be found in Özcan (2012).

Participants

For the behavioral experiments, thirty five simulator pilots (all male) with a moderate to an advanced experience level participated. They were selected through online simulation communities after completing a pre-flight scenario at home. Simulator pilots who also have professional piloting experiences were not accepted in the study[1]. Details of the pre-flight scenario can be found in Özcan (2012). All participants were native speakers of Turkish. Their mean age was 30.7 and their average total flight hour was 1356[2].

Design

The Simulated Flight. In the first task, participants were asked to perform a simulated flight as a pilot for Cessna-172 fixed wing aircraft in Microsoft Flight Simulator 2004 environment. During the flight several uncommon events occurred. Questions about the current status of flight were asked for an assessment of the participant's situation awareness (SA). Similar to the original scenario from Çak (2011), for a cognitively demanding scenario, simulation duration was planned as 75 minutes. The take-off and climb phases were designed to be standard and required a low workload demand for familiarization at the beginning of the experiment. However, after the first 25 minutes, the weather became worse and several equipment failures should have resulted in an increased workload for cruise, descent and approach phases of the flight. The original scenario was retained in terms of novel events experienced during the flight. The novel events were icing, rain, turbulence, crosswind, low visibility, low ceiling and planned failures in the equipment (VSI, RMI/HSI Compass, and ASI). The motive was to introduce high workload and stress, which reveals cognitive differences among participants. Queries were administrated with SPAM technique (Durso and Dattel, 2004) for eight online measurements and with SAGAT technique (Endsley, 1995a) for thirteen offline measurements. The original SPAM technique was modified by removing the "reject to answer" option in order to assure high workload during the SA queries (Çak, 2011). Online queries were carried out orally while offline queries were asked in two sessions, with five and eight questions at a time. Further details of the scenario can be found in Özcan (2012).

AOSPAN. For an assessment of working memory capacity, the Automated Operation Span (AOSPAN) test was used (Unsworth et al., 2005). This test is the computerized version of operation span task (Turner & Engle, 1989) and taps on complex working memory. For this task, three to eight letters are shown in a sequence to be remembered. Participants also have to correctly answer 85 % of mathematical questions asked before each letter. After the eight-letter sequence is recalled, the score is calculated as the sum of perfectly remembered sequences through the task. No ceiling effect was observed despite highly qualified participants. It has been administrated using E-prime software (Psychology Tools Inc.) and details of the procedures for AOSPAN task can be found in Unsworth et al. (2005).

Stroop. To measure inhibition capability in attention, the Stroop task, an indicator of well-managed attention (MacLeod, 1991), was used. The task was based on the ability of inhibiting a habitual response in favor of the goals of the task. Inhibition came into play where participants had to suppress the prepotent response, word reading, in favor of color naming. The difference in response times between congruent and incongruent cases was calculated as the inhibition delay as commonly used in the literature (MacLeod, 1991). In response time calculation, any wrong color namings have been excluded. This task has also been administrated using E-prime software.

Coşkunöz Visual Attention Test. Divided attention was measured by a dual visual task, Coşkunöz Visual Attention Test developed by Er, Sümer, Koku, Mısırlısoy, Coşkan, Erol-Korkmaz, Sümer, Ayvaşık and Eriş (2011) in which participants were expected to follow and respond to two visual tasks running at the same time on the left and the right sides of the screen. On the left side, a red dot traveled through the borders of a hidden shape without leaving a trace. When it was finished, participants were asked to find this shape among five alternatives. On the right side, four drawings of a tool were presented, one of which was slightly different. The different one was expected to be selected. The task on the left side can be considered as the primary task, since, as it runs, the task on the right side (the secondary task) runs for 5 to 7 times. The divided attention capability was measured by the combined score which is the number of the correct answers for the secondary task that were answered in the period where the primary task was correctly answered.

[1] One participant working as a professional Air Traffic Controller and two participants who were student pilots participated in the study. Data from these participants were not found to be outliers.

[2] Total flight time for simulator pilots include piloting experiences from different platforms.

Results

Individual Cognitive Differences and Situation Awareness

Table 1.
Correlations between variables

Variable	1	2	3	4	5	6	7
1. Online SA Score							
2. Online RT	.01						
3. Offline SA Score	.20	-.28					
4. Combined SA Score	**.72***	-.19	**.82***				
5. Working Memory	.01	.24	-.00	.00			
6. Inhibition	-.31	**.50***	**-.55***	**-.57***	.27		
7. Divided Attention	.03	-.10	.12	.10	.16	-.22	
8. Log (Expertise)	.14	-.22	**.49***	**.43***	.08	-.27	-.02

Note. N = 35, * p <.01

Online SA Reaction Time (RT) values have been obtained by summing up RTs for successfully answered online queries. Combined Score was the summation of Offline and Online SA scores. Individual cognitive differences were represented by scores from AOSPAN, Stroop and Coşkunöz visual attention tasks. For expertise, participants' total flight time on several simulation platforms have been used.

For the purpose of finding contributions of predictors to SA measures, expertise, working memory, inhibition and divided attention capacity scores have been used in linear multiple regression analyses. Four regression analyses for Offline SA, Online SA, Online RT and Combined SA have been done on SPSS (Version 20) using data from thirty-five participants. Obtained data has been analyzed for descriptive statistics first. Due to the non-linear relations observed between expertise and dependent variables, log transformation is applied on expertise values.

Correlation analysis has been performed to see the connections between the variables as given in Table 1. Combined scores have significant correlations with Online and Offline Scores since it is calculated as a sum of the two. Online RT, Offline and Combined Scores are significantly correlated with one of the predictors, inhibition capacity and Offline and Combined SA Scores are significantly correlated with expertise. Correlations between SA scores and inhibition capacity are negative since inhibition delay represents the delay in the incongruent cases in Stroop task. Among the predictors, no correlation has been found.

Considering the results from regression analysis, situation awareness is associated with only inhibition capacity and expertise for simulator pilots. Working memory and divided attention capacity were not found to be predictors for any SA measurement.

Comparison with Professional Pilots

As mentioned before, similar tasks from Çak's study with professional pilots have been carried out with simulator pilots and comparable results have been obtained. For professional pilots, 58% of variability in offline SA measures was accounted for by variances in working memory and expertise measures (Çak, 2011). In Çak's analysis, WMC was the most successful predictor (β = .675, t(31)= 5.31, p<.00), whereas the other predictor, expertise (β = .278, t(31)= 2.35, p<.05) was not that successful. For online SA measures (average RTs for correct answers in online queries), 52% of variability was accounted for by variances in inhibition, divided attention and expertise measures. The predictors in order of strength are listed as expertise (β= -.470, t(31)= -3.73, p<.001), divided attention (β= .313, t(31)= 2.25, p<.05) and inhibition (β = .260, t(31)= 2.058, p<.05). However, for simulator pilots inhibition capacity and expertise were found to predict online and combined measures of SA while working memory and divided attention capacities were not predictive. Further results from the regression analyses can be found in Table 2. In comparison to the preceding study carried with professional pilots (Çak, 2011), results from the current study did not reflect a clear picture of a cognitive grounding for SA especially for individual cognitive differences in working memory and divided attention capacities.

Table 2.
Regression Results for SA measurements.

		Predictors			
Dependent Variable	Regression Result	Inhibition	Log (Expertise)	Working Memory	Divided Attention
1. Online SA Score	Not successful	-	-	-	-
2. Online RT	adjusted R^2 of .17 (F(4,34) = 2.756, p= .05)	β = .434, t(30)= 2.481, p< .05	-	-	-
3. Offline SA Score	adjusted R^2 of .38 (F(4,34) = 6.233, p< .01)	β = -.495, t(30)= -3.270, p< .005	β = .372, t(30)= 2.649, p< .05	-	-
4. Combined SA Score	adjusted R^2 of .36 (F(4,34) = 5.732, p< .005)	β= -.545, t(30)= -3.536, p< .05	β = .297, t(30)= 2.077, p< .05	-	-

Note. N = 35

Obtained results from the two groups were compared using independent group t-tests. Results showed that there are no significant differences between groups except for the Offline SA scores. Offline SA scores from professional pilots (M=720.37, SD=195.8) and simulator pilots (M=602.86, SD=166.6) were significantly different from each other; t(68)=2.704, p<.01. Further details of the statistical analysis can be found in Özcan (2012). Despite the similarities in cognitive capacity tests and expertise, professional pilots were distinctively more successful in offline SA queries[3]. The differences in pilot training backgrounds and practice systems structures are candidate reasons to explain this finding.

Discussion

Considering the whole regression results obtained for simulator pilots, first unexpected finding is the absence of expertise as a consistent predictor. The regression results report that expertise contribute to the explanation of the variances in only Offline SA and Combined SA scores. Leaving the theoretical problems of measuring SA aside, the first reason is possible errors in assessment of expertise. In Çak's studies, professional pilots' expertise was determined as the number of flight time spent in the specific full flight simulator, not of the actual flight. For simulator pilots, their total simulation flight time from different platforms were used due to to the variety in their simulator experiences. As a result, this assessment technique for expertise comes with precautions for its contribution to the prediction of SA measurements. Another important issue for this study is that WMC is found to be not explaining any of the variances in the SA measurements. Working memory is considered to have a central importance for SA (Durso & Gronlund, 1999; Endsley, 1995b). Flight critical tasks, systems and timely information are kept and processed by working memory (Wicken, 1999). Also, there are several studies in which the correlation between SA measures and WMC is given (Durso et al., 2006; Gonzales and Wimisberg, 2007). However, in this study, no correlations or similarities in variances has been observed between WMC and SA measures. A possible explanation to this finding is the vast range of differences in simulator piloting practices

[3] Online SA scores between professional and simulator pilots could not have been compared since in Çak's study only online SA reaction times were measured.

compared to professional piloting. Simulator pilots are generally self-educated and have their own unique ways of piloting due to the lack of formal education. During the experiments, it is observed that they learn simulator piloting with the help of autopilots and automatic navigation devices. Overall considerations on the results point that differences between professional and simulator pilots are more foundational than expected and simulator pilots' performances on SA measurements are not determined by the systematical factor of working memory capacity, but possibly determined by individual self-training differences.

Following the discussion on lack of a formal education, the results for the rest of predictors, inhibition and divided attention seem to lose their importance. However, even under these considerations, it is important to note that inhibition capacity happened to be the consistent predictor for Online RT, Offline and Combined SA measurements. Attention control capability captured by Stroop task, unlike utilization of working memory for elements of flight, is found to be effective in SA performance. Since the SA measurements were carried out as the participants were busy with piloting, answering these queries required a successful management of attention. Possibly with this connection, inhibition capacity turned out to be a consistent predictor. Along similar lines, the reason why divided attention did not turn out to be a good predictor can be explained. Compared to the real flight situation, simulation environment is simple in terms of environmental factors. Simulation environment consisted of a PC and a joystick while real flight contains two environments, the cockpit and the outside of the airplane. Professional pilots observe both the equipments inside the plane and weather conditions outside the plane. Consequently, it might be proposed that divided attention capacity for simulator pilots is not as important as it is for professional pilots. Nevertheless, due to the effects of unsystematic differences in simulator pilots' practices as mentioned above, these comments have to be interpreted with caution.

Acknowledgements

Authors gratefully acknowledge contributions from Dr. Serkan Çak, Dr. Mine Mısırlısoy, Dr. Canan Sümer, her team in METU Psychology Department and the Coşkunöz Metal Forming Company. Authors also thank THK Antalya Branch members for their organizational support and simulator pilots for their participation. This paper is based on the Master Thesis of the first author (Özcan, 2012).

References

Breton, R., & Rousseau, R. (2001). *Situation awareness: A review of the concept and its measurement* (Technical Report No. 2001-220). Valcartier, Canada: Defense Research and Development.

Carretta, T. S., Perry Jr., D. C., & Ree, M. J. (1996). Prediction of situational awareness in F-15 pilots. *The International Journal of Aviation Psychology,* 6(1), 21-41.

Çak, S. (2011). *Effects of working memory, atention, and expertise on pilots' situation awareness.* Unpublished doctoral dissertation, the Middle East Technical University, Ankara.

Durso, F. T., Bleckley, M. K., & Dattel, A. R. (2006). Does situation awareness add to the validity of cognitive tests? *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 48(4), 721-733.

Durso, F. T. and Dattel, A. R. (2004). SPAM: The real-time assessment of SA. In S. T. S. Banbury, editor, *A Cognitive Approach to Situation Awareness: Theory, Measures and Application* (pp.137–154). Ashgate.

Durso, F. T., & Gronlund, S. D. (1999). Situation awareness. In F. T. Durso, R. Nickerson, R. Schvaneveldt, S. Dumais, S. Lindsay, & M. Chi (Ed.), *The handbook of applied cognition* (pp.283–314). Wiley.

Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 37(1), 65–84.

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 37(33), 32-64.

Endsley, M. (2006). Expertise and situation awareness. In K. A. Ericsson, N. Charness, P. Feltovich, & R. Homan, (Eds.), *The Cambridge Handbook of Expertise and Expert Performance,* (pp.636–651). New York: Cambridge University Press.

Endsley, M. (1997). The role of situation awareness in naturalistic decision making. In C. Zsambok and G. Klein, (Eds.), *Naturalistic Decision Making*, (pp.269-283). Lawrence Erlbaum Associates.

Endsley, M. R., & Bolstad, C. A. (1994). Individual differences in pilot situation awareness. *International Journal of Aviation Psychology,* 4(3), 241-264.

Er, N., Sümer, H. C., Koku, B., Mısırlısoy, M., Coşkan, C., Erol-Korkmaz, H. T., Sümer, N., Ayvaşık, H. B., & Eriş, A. (2011). *Çoşkunöz visual attention test.* Unpublished instrument.

Gonzalez, C., & Wimisberg, J. (2007). Situation awareness in dynamic decision making: Effects of practice and working memory. *Journal of Cognitive Engineering and Decision Making,* 1(1), 56-74.

Horswill, M. S., & McKenna, F. P. (2004). Drivers' hazard perception ability: Situation awareness on the road. *A cognitive approach to situation awareness: Theory and application*, 155-175.

Johannsdottir, K. R. (2004). Situation awareness and working memory: *An integration of an applied concept with a cognitive fundamental process.* Unpublished doctoral dissertation, Carleton University, Ottowa.

Macleod, C. (1991). Half a century of research on the stroop effect: An integrative review. *Psychological Bulletin,* 109, 163-203.

McCarley, J. S., Wickens, C. D., Goh, J., & Horrey, W. J. (2002). A computational model of Attention/Situation awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 46(17), 1669-1673.

Özcan, O. O. (2012). *Exploring the Effects of Working Memory Capacity, Attention, and Expertise on Situation Awareness in a Flight Simulation Environment.* Unpublished master's thesis, the Middle East Technical University, Ankara. Retrieved from https://dl.dropbox.com/u/14641303/Ozcan_MSc.pdf.

Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *The International Journal of Aviation Psychology,* 1(1), 45-57.

Sohn, Y. W., & Doane, S. M. (2004). Memory processes of flight situation awareness: Interactive roles of working memory capacity, long-term working memory, and expertise. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 46(3), 461-475.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language,* 28(2), 127-154.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Enle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods,* 37(3), 498-505.

Wickens, C. D. (1999). Cognitive factors in aviation. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay, & M. T. H. Chi (Eds.), *Handbook of Applied Cognition* (pp. 247–282). Wiley.

SITUATION AWARENESS AND SITUATION ASSESSMENT: HOW ARE THEY RELATED?

Norman R. Hertz
Progeny Systems Corporation
Manassas, VA 20110
Michael T. Hertz
Progeny Systems Corporation
Manassas, VA 20110

The goals of human factors practitioners and industrial psychologists in situation awareness milieu are to assess individuals, teams, and organizations to measure current performance and predict future performance. The value in assessing situation awareness (SA) is that it impacts performance. Endsley (1995a) stated "SA provides the primary basis for subsequent decision making and performance in complex, dynamic systems." The purpose of this paper is not to discuss or propose another means of describing or measuring SA but instead to propose the role of situation assessment (SAS) in identifying the nature of SA, product and/or process, regardless of the definition or explanation of SA. It appears the research efforts bypassed the fundamental building blocks and rationale for describing a concept and instead developed measuring instruments and explanations of a complex phenomenon. Simply stated, SA and SAS are related by SAS providing the mechanism-of-action for the measurement of SA.

Situation awareness (SA) is passive when thought of as only a product. It doesn't imply action, making decisions, or doing anything—it just means taking in information and being aware of the variables impinging on the situation. On the other hand, situation assessment (SAS) when thought of as a process implies actions and decision making. This perception very quickly leads into a discussion of whether situation awareness is process or product or both. One of the problems with the theoretical construct of situation awareness is that it appears intuitive and requires only common sense to understand its implications so it must really exist. The opposite problem is that SA is obscure and difficult to define and measure so that even scientist practitioners do not agree. Situation awareness is definitely accepted in the scientific community as a concept or construct but other than that there is little agreement about how one goes about measuring it. There is some agreement by the scientist practitioners, although not totally, that SA is dynamic and it is both a product and a process. SA is the process of developing awareness of a situation and the product of awareness that is developed. SA should be viewed in terms of the process involved in its development and in the end product of what it comprises (Salmon, Stanton, Walker and Jenkins, 2009).

Because of the differences in approaches to measuring SA, the focus of this paper is to advance the study of SA by identifying measures that would predict SA regardless of its definition. The most often applied methodology for measuring situation awareness (SA) is the Situation Awareness Global Assessment Technique (SAGAT) model proposed by Endsley (1995b). She (1995b, page 36) defined situation awareness as "the perception of 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." A basic definition of situation awareness is that it refers to up-to-the-minute cognizance necessary to operate or maintain a system. However, situation awareness must be created based upon information obtained from many sources in the environment in conjunction with the person's schemata. The information in the environment does not enter into the person's consciousness without effort and when it does, it is processed by the person's filter. The quality and accuracy of situation awareness depends upon the ability of the person to adequately assess and process all the available information.

The interest is not in determining whether the individual was aware of the situation but it is more important to know if the person performed an accurate assessment of the situation in order to take appropriate action. By focusing on situation assessment, then we are able to provide interventions to individuals and teams to build relevant schema. Then in unfamiliar situations, the individuals or teams do not have to depend solely upon working memory to respond but instead are able to assess schema. Using this line of reasoning, training could be presented with realistic situations that would be stored in long-term memory to be recalled when similar situations are encountered.

The statement that the loss of situation awareness (SA) is the leading cause of human error in military aviation mishaps cited by Flach (1995 p. 151) appears to be circular reasoning: How does one know that SA was lost? …because the human responded inappropriately. Why did the human respond inappropriately? …because SA was lost. Patrick and Morgan (2010) stated that because too little attention was given to Flack's warning, "Situation Awareness: Proceed with Caution," there is little consensus and different theoretical and methodological approaches have been used to study SA.

Another approach to defining situation awareness is not needed. Salmon, Stanton, Walker and Jenkins (2009) conducted a review to identify and understand the different SA measures presented in the literature. They concluded that the SA literature is quite disparate and that many models exist but they present the construct quite differently (p. 32). Salmon, et al. (2009) believed that the perceptual cycle model of Smith and Hancock (1995) offers a complete description of how SA is achieved and maintained. Salmon, et al. (2009) indicate that the perceptual model of Adams, Tenney, & Pew (1995) was worth further study for the same reasons as the Smith and Hancock model. The models include the notion of a continuous cycle SA acquisition and maintenance including both the process and the product. They believed that the perceptual cycle model comes closest to most accurately describing the construct but has not received sufficient attention.

Salmon, et al. (2009) presented results of their review in Table 3.2, pages 52-55, "Summary of SA measurement techniques review." They identified 18 approaches. Other than with Situation Awareness Global Assessment Technique (SAGAT) from Endsley (1995a) and Situation Awareness Rating Technique (SART), Salmon, et al, 2009, p 56 concluded there was very limited validation evidence associated with the SA measurement techniques . A review of the literature could easily locate hundreds of articles and scores of books discussing various aspects of situation awareness for individuals and teams. It is as if situation awareness were examined from another perspective that the "truth" could be discovered. Certainly, there is a situation awareness construct but the construct would be most useful if it were able to predict behavior and job performance. Tenney, Adams, Pew, Huggins, and Rogers (1992) state that situation awareness contributes to good performance but is not synonymous with it. Vaitkunas-Kalita, Landry, and Yoo (2011) states "SA clearly appears to be an independent phenomenon that influences behavior."

The statement by Sarter and Woods (1991), "Situation awareness is based on the integration of knowledge resulting from recurrent situation assessments" is the impetus for proposing a model that supports the belief that SAS can be used to predict SA across dynamic settings. SA depends upon the ability of individuals to constantly update their conscious and mental resources. This paper will propose SAS constructs that provide the necessary information to maintain a current SA.

SITUATION ASSESSMENT

A methodology for measuring the predictors (elements) involved in the development of situation awareness is needed. Situation awareness is a hypothetical construct that cannot be measured directly. Fracker (1988, page 103) states that "Defining situation awareness determines what is to be measured but does not suggest how it should be measured. For this latter purpose, a model of situation assessment is needed."

However, the processes of situation assessment that are used to develop situation awareness can be measured directly. Situation assessment is viewed as a process that refers to the various perceptual and cognitive activities involved in constructing, updating, and revising the state of awareness (Adams, Tenney & Pew, 1995). Sarter and Woods (1991) state that situation awareness is based on the integration of knowledge resulting from recurrent situation assessments. Endsley (1995 b, page 36) recognized the importance of SAS in referring to SAS as "the process of achieving, acquiring, or maintaining SA." Therefore, SAS affects SA as an unitary concept, neither simply a process or product solely but as integrated concept. The focus of this paper is not to link SAS to either process or product of SA but to link SAS to the concept of SA. The application of SAS will be to measure and predict SA which is constantly being updated. SAS will be able to able to measure SA without active interventions.

Situation assessment implies that the individual has responsibility for processing and understanding the forces acting within the situation and the consequences of action. Situation assessment is trainable and occurs prior to an event. The advantage of a situation assessment approach is that it does not depend solely on native intelligence but is trainable by building schemata.

Given that there is no universally agreed upon definition of SA, the direction that will be proposed in this paper is to examine the role of situation assessment (SAS) variables that will generalize to most of the approaches defining and describing situation awareness. The emphasis will be on SAS variables that are observable, measureable, and testable. We propose to identify the variables of SAS that will predict SA. Also, we propose to identify measurement tools that would be used to identify the SAS variables that affect SA. We are applying the beliefs as stated by others that SAS variables are essential to the development of SA. As long as SA is considered as unobservable phenomena and not directly measureable, research will continue. SA can be thought of as the full employment act for human factors professionals.

The research for the identification of SAS variables will apply a construct validation approach. A construct validation is appropriate because the SAS variables are measureable. The validated SAS variables can predict SA both as a process and as a product. The approach taken to validate the situation assessment variables will be modeled after the Campbell and Fiske (1959) approach. The SAS variables will be validated with the multitrait-multimethod approach to identify the variables that are related to the prediction of SA. Instead of developing another tool that attempts to measure SA directly, the SAS approach will discover the underlying constructs that are predictive of SA regardless of the changes in the environment and constructs that are possessed by individuals.

Context-free Situation Assessment Elements

To successfully develop a means to measure SA, the model of situation assessment should be context free. A context-free model means that SAS is independent of the environment in which SA is operating; therefore, SA is also context free. If it were not, then SA is dependent upon the specific conditions and an individual's level of SA could not generalize across the setting. Because SA is viewed as dynamic; each new event would require reassessing SA. Proposed elements of situation assessment are listed in Table 1. The first step in developing measurement instruments is to develop operational definitions for the variables. The process of developing operational definition as stated by Kerlinger (1973, p 31) is to assign meaning to a construct or variable by specifying the activities or "operations" necessary to measure it. Further explanation is that it gives meaning to a variable by explaining the actions to be taken by the investigator to measure the variable.

Smith and Hancock (1995) defined SA as externally directed consciousness because it is not until the externally directed task is made explicit that the observed behavior achieves the status reserved for SA. (Hauland, page 290). This interpretation assists in supporting a methodology for applying the development of behaviors for the purpose of measuring SAS. To develop the behaviors for the elements listed in Table 1, the critical incident methodology will be applied (Flanagan, 1954). In applying the critical incident methodology, subject matter experts will identify behaviors that represent effective levels of performance for each of the elements. In order to ensure that the behavioral statements are at the same level of effectiveness (a common metric), the behavioral statements will be scaled using the methodology proposed by Smith and Kendall (1963). The behavioral rating system for measuring assessment will be a matter of combining the ratings for each the elements and then summing across the elements.

The approach proposed is designed to be suitable for both individual and team situation assessment. By approaching SAS from an element level of measurement, interventions could be designed to provide training in the areas of weakness. It is not a reasonable process to wait until an individual or team fails to exhibit appropriate situation awareness before designing a training program. By taking a behavioral indicator approach to SAS, the individual or team SA potential can be identified based on the assessments developed for each element. Situation assessment is composed of these elements. The higher the score on the situation assessment elements will predict better SA. The method of scoring is very similar to the SAGAT system which arrives at a score for one of the three levels by adding the score based on the probes.

The advantage of approaching the measurement of SA from the SAS perspective is that an individual's SA aptitude can be determined prior to allocating job-related training resources. Not all individuals are equally capable of performing dynamic work or maintaining SA so the better process would be to identify individuals who are more likely to be successful prior to placing them in jobs.

Table 1. Hypothesized situation assessment elements (behavioral and cognitive).

Abstract reasoning	Ability to draw meaning from events and behaviors that appear unrelated to current endeavor
Attention	Cognizant of the activities in the environment
Automaticity	Ability to use past knowledge and experience with minimum expenditure of mental energy
Dynamics	Ability to adjust to new and novel situations
Encoding skill	Ability to interpret the influence of activities on the outcome —the skill of assigning meaning to events appearing abstract
Five-factor model of personality traits	Conceptualizes personality in terms of five basic dimensions: • Extraversion • Agreeableness • Conscientiousness • Neuroticism • Openness to experience

Mapping	Ability to fit new patterns of behavior within previously developed schema
Metacognition	Awareness or analysis of one's own learning or thinking processes
Motivation	Demonstration of sufficient engagement to ensure success
Multitasking	Ability to attend to multiple and competing activities
Pattern recognition	Ability to identify similar or dissimilar patterns of behavior
Perception	Details of the events are recognized and processed
Prediction	Ability to recognize the short- and long-term effects of actions or inactions
Psychophysiological variables	• EEG • EKG • Eye blink
Spatial ability	Ability to view stimuli in multiple dimensions

The selection of the situation assessment variables was based on applied psychology and as suggested by human factors literature. The listed elements are not an exhaustive list and it is reasonable that during the proposed study, subject matter experts will add or remove elements. Essentially the study of situation awareness is to predict performance such as being able to safely pilot an airplane or pilot a ship when there are multiple stimuli, competing demands, and interruptions. Some of the elements were selected for the purpose of identifying those that are amenable to training. Some elements are similar to personality traits in that they are enduring and not subject to change. The advantage of using the behavioral indicators as a means of measuring the elements is that some elements are not observable and can be measured only when they are represented as behaviors.

ACKNOWLEDGEMENTS

The opinions and cited research in this paper do not reflect the views of Progeny Systems Corporation.

REFERENCES

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors, 37*, 85-104. doi: 10.1518/001872095779049462

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminat validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Endsley, M.R. (1995a). Measurement of situation awareness in dynamic situations. *Human Factors, 37*, 65-84. doi: 10.1518/001872095779049499

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32-64. doi: 10.1518/00187209577904953

Flach, J. M. (1995). Situation awareness: Proceed with caution. *Human Factors, 37*, 149-157. doi: 10.1518/001872095779049480

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327-358.

Fracker, M. L. (1988). A theory of situation assessment: Implications for measuring situation awareness. In *Proceedings of the Human Factors and Ergonomics Society 32nd Annual Meeting* (pp. 102-106). Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193128803200222

Hauland, G. (2008). Measuring individual and team situation awareness during planning tasks in the training of en route air traffic control. *International Journal of Aviation Psychology, 18*, 290-304. doi: 10.1080/10508410802073333.

Kerlinger, F. N. (1973). Foundations of behavioral research (2nd ed.). New York, NY: Holt, Rinehart, and Winston, Inc.

Patrick, J., & Morgan, P. L. (2010). Approaches to understanding, analysing and developing situation awareness. *Theoretical Issues in Ergonomics Science, 11*, 41-57. doi: 10.1080/14639220903009946

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, G. P. (2009). *Distributed situation awareness: Theory, measurement and application to teamwork*. Burlington, VT: Ashgate Publishing Company.

Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *International Journal of Aviation Psychology, 1*, 45-57. doi: 10.1207s15327108ijap01001_4

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: an approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47*, 149-155.

Smith, K. & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. Human Factors, 37, 137-148. doi: 10.1518/001872095779049444.

Tenney, Y. J., Adams, M. J., Pew, R. W., Huggins, A. W. F., & Rogers, W. H. (1992). *A principled approach to the measurement of situation awareness in commercial aviation* (NASA Contractor Report 4451). Langley, VA: NASA.

Vaitkunas-Kalita, S., Landry, S. J., & Yoo, H. –S. (2011). Coincidence between the scientific and folk uses of the term "situation(al)" awareness in aviation incident reports. *Journal of Cognitive Engineering and Decision Making, 5*, 378-400. doi: 10.1177/1555343411424694

ASSESSING THE EFFICACY OF SITUATION AWARENESS PROBE QUESTIONS FOR PREDICTING AIR-TRAFFIC-MANAGEMENT PERFORMANCE

James D. Miles, Kim-Phuong L. Vu, Dan Chiappe, Thomas Strybel
Center for Human Factors in Advanced Aeronautics Technologies,
California State University Long Beach
Long Beach, CA USA
Vernol Battiste
NASA Ames Research Center, San Jose State University Foundation
Moffett Field, CA USA

We conducted an exploratory data analysis as a step toward modeling components of SA. It was based on data collected from situation awareness probe questions that were used in large-scale air traffic control simulations over 5 semesters of an ATC radar internship in the Center for Human Factors in Advanced Aeronautics Technologies (CHAAT). Three components of SA were generated by a principal component analysis that we label "Action Relevant," "Distance Relations," and "Low Priority." The analyses provide a data-driven scheme for categorizing probes related to SA that can be used in future evaluations of NextGen concepts and technologies.

Situation awareness (SA) is central to evaluating future airspace systems being considered by the FAA for implementation in the Next Generation Air Transportation System (NextGen). SA refers to the operator's understanding of the evolving situation that he or she is in for the purpose of projecting system states in the near future. The specifics of this definition have been debated in many recent reviews (e.g., Salmon, Stanton, Walker, and Jenkins, 2010; Chiappe, Strybel, and Vu, 2012; Jeannot, Kelly, and Thompson, 2003). Although airspace operators report having clear ideas of what SA means (e.g., D'Arcy and Rocco, 2001), Chiappe et al. (2012) showed that diverse theoretical perspectives on SA and its measurement are grounded in very different approaches to cognitive science. Endsley (1995a), for example, assumes that processing and representation all takes place in the conscious mind of operators. Recent distributed or situated conceptions of SA are instead based on a view of cognition that assumes operators offload task representation and computation to the external environment in order to limit use of internal processing resources (Salmon et al., 2010; Chiappe et al., 2012). According to situated SA theories, operators create partial representations of a situation that are constantly updated, and often internally store where to find information in the environment, rather than storing the information itself. SA, in this view, therefore exists in the interaction between the operator and his or her task environment.

These diverse concepts of SA have implications for how the construct is measured. Probe techniques are promising because they can assess an operator's awareness of specific information needed for adequate performance. Probe queries can be administered either offline or online. Endsley's Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1995b) is an offline method. In SAGAT, the operator is queried about task information when a scenario is frozen and the displays are blanked. SA is measured by the number of correct answers. This technique is consistent with the notion that knowledge of the task environment is stored internally because SA is determined by probing only the operator's working memory for task information. Evidence for the criterion validity of SAGAT has been reported (e.g., Endsley, 2000; Endsley, 1990b; Gronlund et al., 1998), and SAGAT has been used in a variety of settings such as air traffic control, aviation, and nuclear power plant operations. It is the most widely-used probe technique for situation awareness assessment to date. Online probe techniques such as Durso et al.'s (2004) Situation Present Assessment Method (SPAM) are consistent with a situated approach to SA. Operators are queried about task information in real time while performing their tasks with all displays and controls available for answering probe queries. Both, the number of correct responses and the latency of responses to queries are assumed to be indicators of SA, but only response latencies should be sensitive to whether information is offloaded to the environment. Evidence for the validity of this online probe technique has accumulated in recent years (e.g., Durso et al., 2004; 2006; Bacon et al., 2011; Strybel et al., 2013).

An important factor determining the effectiveness of all probe methods is the information contained in questions themselves. Yet, this factor has received little attention in the literature. For both offline and online probe techniques, it is recommended that the queries be developed in consultation with subject matter experts. Endsley et al. (2000) recommends that SAGAT queries be developed from a systematic Goal Based Task Analysis in which task goals, required information, and required SA for meeting the goals are identified. Durso et al. (2004) also recommend that queries be developed in conjunction with subject matter experts, but did not advocate a formal process of question development. However, for both probe methods, SA probe queries have been specific to the scenarios and variables manipulated, making it difficult to compare changes in SA across different simulations, experiments and operating concepts.

Work in our lab has been focused on developing and standardizing categories of probe-questions for assessing SA in current and future airspaces so that changes in awareness of specific information can be determined. We have examined several probe category schemes, but each has been found less than adequate for the purposes of comparison. Initially, we developed probe categories based on level of processing (recall, comprehension) and time frame (past, present, future) consistent with Endsley's conception of SA. Dao et al., (2009), for example, determined that pilot probe latencies for future questions were significantly longer than latencies to questions asking about present or past task information, suggesting that pilots were less aware of future information, especially for recall questions. Dao et al. noted that these results could be an artifact of the part-task simulation, however: pilots were queried about their airspace following individual conflict-resolution trials, and at that point the traffic in the scenario was frozen. Strybel et al. (2009) used the same information processing categories to measure ATCos' SA. They showed that categories based on task-specific information, such as conflicts, were more predictive of performance than information processing level or time frame. Subsequently, we developed probe categories based on the information required for different task components and operators; for example, conflicts, sector/aircraft status, command /communications, traffic and weather (e.g., Bacon et al., 2011Strybel et al., 2013). These categories were developed prior to a simulation based on consultations with subject matter experts and were more successful in predicting operator performance than the queries based on information processing categories. Recently, Morgan et al. (2012) categorized probe questions as part of a test of the situated SA approach. Morgan et al. hypothesized that questions based on high-priority task information would be answered more quickly than low-priority information, and that general-task questions would be answered more quickly than task-specific queries. This is because general information and high priority information should be stored in the head. Morgan et al. determined that probe latencies were faster for general information than for specific information, consistent with a situated situation awareness approach. However, they did not find an effect of task priority on probe latencies.

In summary, probe techniques are promising methods of assessing SA, yet their usefulness in comparing changes in SA across different simulations and comparing changes brought about by different concepts of operation are limited by a lack of standardized probe categories. Previous categorization schemes have been based on either theoretical assumptions of SA, or information relevant to specific task components. The problem of categorization is difficult because theory-based categories and task-specific categories are not mutually exclusive. For example, high-priority queries used in Morgan et al. (2012) were mostly questions on conflicts because safety is the highest priority of air traffic controllers. Similarly, future-oriented questions used in Strybel et al (2009) can be classified as conflict questions because detecting conflicts requires projection into the future.

In the current paper, we conducted an exploratory data analysis as the first step toward modeling components of SA. The analysis was based on data collected from probe questions related to ATC situational awareness that were used in simulations over 5 semesters of an air-traffic-control radar internship in the Center for Human Factors in Advanced Aeronautics Technologies (CHAAT). The advantage of such simulations is that they allow for the collection of real-time SA in a realistic ATC setting. In the current paper, we combined data from interns in these 5 simulations, which provided an appropriate N for conducting a principle component analysis.

Method

Participants

Data for this analysis were taken from 71 students enrolled in a radar internship course over five semesters (approximately 13 per semester) between 2010 and 2012. All students were enrolled in an FAA Collegiate Training Initiative (CTI) at the time of the course and had taken some courses in aviation sciences on FAA regulations and air traffic control operations. Students had little-or-no radar experience, however.

Training

The 16-week radar internship was designed to provide students training and practice managing traffic in a simulated en-route sector (ZID-91), using both current day manual skills and potential NextGen tools. The simulation software was Multi Aircraft Simulation System (MACS; Prevot et al., 2002). Students completed a 3-hour lab period and 2-hour lecture period each week. A retired, full performance Level (FPL) ATC taught the internship. Students learned to manage traffic with current day ATM techniques such as altitude, speed, vectoring, and structure, and current-day basic ATM phraseology. They also learned to manage traffic with simulated NextGen tools, specifically, integrated Data Comm, a conflict probe tool and a trial planner. The sequence in which current day and NextGen tools were introduced, as well as scenarios containing different mixtures of equipped and unequipped aircraft, varied somewhat between different semesters. At the end of each semester, however, all students had roughly equivalent practice with manual and NextGen skills, and had managed scenarios having 0% to 100% equipped aircraft.

Situation Awareness Assessment

At the midterm and end of the course, students were tested on three 40-50 minute scenarios differing in the percentage of equipped aircraft. These tests were intended to determine the effects of different training approaches, and were not aimed at evaluating their performance for classroom assessment (the instructor never saw the results). In the first semester, the equipage levels were 50% overall and varied within the scenarios between 25% and 75%. In the last four semesters, equipage levels were 0%, 50% and 100%. SA was also measured during these tests using an online probe technique. Students were instructed to respond to online probe questions about their airspace that were presented every 3 minutes beginning 4 minutes into the scenario. Probe questions were administered on an adjacent computer with touch-input. Each question began with a "ready for question" prompt, and the participant was instructed to respond affirmatively only when their workload would allow it. When the ready prompt was accepted an SA question was presented on the touch screen. The participant selected the answer by touching one of the alternatives shown. If the ready prompt was not accepted after one minute, it was withdrawn and presented again after two minutes, thus preserving the 3- minute interval.

Probe Question Development

For each scenario, 12-16 probe questions were presented based on scenario length. Four of the probe questions asked for workload ratings, leaving 8-12 SA questions per scenario. The probe question categories used in each semester were conflicts and sector status. Conflicts ask about existing or potential conflicts, including information relative to conflict detection. Status queries asked about current traffic, equipage mixtures, commands and communications, etc. (see examples shown in Appendix A). Probe questions were developed as follows (for more detail, see Strybel et al.,2011): First, we developed counterbalancing schemes so that all probe categories were presented equally often during the scenarios. Then we developed probe "stem" questions, in consultation with subject matter experts. These questions were general, and could be asked at any time in the scenario, or specific, requiring additional information such as aircraft callsigns based on time in the scenario. The wording of questions was vetted with subject matter experts. Once stem questions were finalized, the specific question was inserted into the appropriate time slot and additional detail (i.e., aircraft call sign or waypoint name) was added if necessary. Probe questions were scored by comparing the answers to the questions with a recording of the airspace at the time of the question. All questions were scored by at least two researchers independently. If disagreement was found a third researcher independently reviewed the question and reconciled the answer.

Results

Data Screening and Formatting

Only RTs from correctly answered probe questions (71.3% of all probes) were included for further analysis. Additionally, any probe questions with RTs greater than 40 seconds were considered outliers and exclude from the analysis. The remaining questions were placed in one of nine categories (see Appendix A). Categories were created a priori based on the existing probe question data to best represent general aspects of performance in the ATC simulations. Several such sets of categories were created, each containing between 8-12 categories. These category sets were each subjected to PCA and the set shown in Table 1 was determined to have the most variance explained with the fewest number of components as well as the clearest loadings on the resulting components.

Factor Analysis

A Pearson's correlation including all categories found that all 9 categories correlated at least .3 with at least one other item, and 6 categories had correlations of at least .4 with 1 other item, suggesting moderately reasonable factorability. Since some interns did not receive questions within a particular category, or incorrectly answered probe questions within a category, 32 interns were excluded by the listwise comparison that was used because they had missing data for at least 1 category. This left 39 participants in the analysis. Additionally, the Kaiser-Meyer-Olkin measure of sampling adequacy was .61, which is above the commonly accepted cutoff point of .6. Bartlett's test of sphericity was also significant (χ^2 (36) = 74.98, $p < .05$). Finally, the communalities were all above .4, further confirming that each item shared some common variance with other items. Given these overall indicators, factor analysis was conducted with all 9 items. A principle-components factor analysis using Varimax rotation and listwise comparisons was conducted, with three factors with Eigenvalues above 1 (2.64, 1.63, and 1.36, respectively) explaining 63% of the variance.

The component loading matrix for this final solution is presented in Table 1. We designated the component with loadings from probe questions related to AC separation, Global AC traffic conditions, General amount of upcoming ATC actions, and Next ATC action as "ATC Relevant" information. A second component has clear loadings with Lateral AC Distances, Altitude Distances, and Future Lateral Distances and was therefore designated "Distance Relations" information. A third component had strong loadings on probe questions related to AC Entering the Sector and Memory for Prior Conflicts. Both categories refer to probe questions unrelated to events in the sector at the time of the question, the component was therefore designated "Low Priority" information.

Composite scores were created for each of the three factors by calculating the mean for all questions from categories which had their primary loadings on each factor. Descriptive statistics are presented in Table 2. Response Times (RTs) were longest for Distance Relations probe questions followed by Action Relevant questions. Low Priority probe questions had the fastest RTs – this was likely due to the simplicity of these probe questions, which did not required a comparison or distance judgment of any type. As expected for RTs, data was positively skewed in each component, but the amount of skew ranged from very good to acceptable levels. Kurtosis was also in the acceptable range for all components.

Table 1.
*Component loadings, proportion of variance, and communalities based on PCA with Varimax rotation for 9 items from probes (*N *= 39)*

	Action Relevant	Distance Relations	Low Priority	Communality
Closest AC separation	.79			.66
Global AC Traffic conditions	.77			.70
General amount of upcoming ATC actions	.65			.52
Next ATC action	.55			.45
Lateral distance from AC to waypoint/AC		.82		.73
Altitude distances between AC		.82		.75
Future lateral distances from AC to waypoint/AC		.57		.47
AC Entering the sector			.82	.70
Memory for prior conflicts			.80	.66
Proportion of Variance	.243	.221	.161	.625

Note. Factor loadings < .4 are suppressed

Table 2.
*Descriptive statistics of responses times (RTs) in seconds to probe questions in each category component (*N *= 39)*

	Num of Items	*M* (*SD*)	Skewness	Kurtosis
Action Relevant	4	12 (3.1)	1.08	1.8
Distance Relations	3	14 (3.2)	.66	.91
Low priority	2	9 (3.3)	.66	-.22

Overall, these analyses indicated that three distinct components were underlying probe questions RTs. An approximately normal distribution was evident for the composite score data in the current study. The data were therefore well-suited for parametric statistical analyses.

Discussion

The current analyses provide a data-driven scheme for categorizing probe questions related to SA that can be used in future evaluations of NextGen concepts and technologies. The components of situational awareness generated by the PCA, Action Relevant, Distance Relations, and Low Priority bear some resemblance to existing models of SA. In particular, Probe questions loading on Action Relevant component were related to events such as existing or possible conflicts that must be kept in a state of action-readiness. The Distance Relation component loaded with probe questions that were specific to judgments of distances between items within the sector that ATCs were watching, but these items were not actionable and may not be actively stored in the mind but rather, referenced from the display when needed. These questions would take longer to answer because it would require a visual search, and indeed Distance Relation questions had the highest RTs. Together, Action Relevant and Distance Relation components are consistent with a situated SA approach in which information is represented in the head and in the operator's task environment respectively (Salmon et al., 2010; Chiappe et al., 2012). Lastly, Low Priority probe questions involve information that is either spatially or temporally outside of the current sector.

Although the number of participants in the analysis was relatively low and data was collapsed across several significant manipulations such as equipage levels and training period, we were able to extract clear components related to situational awareness. The next step will be to use this PCA and more formal data modeling to further refine SA probe question techniques. For example, it may be a more effective use of time to exclude Low Priority questions when gauging SA because these questions explain the smallest proportion of variance and do not provide a clear theoretical component of SA. Additionally, the components will be used to predict performance within the ATC simulations, including conflict resolutions and losses of separation (LOS). Ultimately, a data-driven model of SA will further resolve discrepancies between theory-based models and accommodate additional factors such as workload and ATC expertise.

References

Bacon. L.P., Strybel, T.Z., Vu, K-P., L., Kraut, J. M., Nguyen, J., Battiste, V. and Johnson, W. (2011). Situation awareness, workload, and performance in midterm NEXTGEN: effect of variations in aircraft equipage levels between scenarios. *Proceedings of the 16th International Symposium on Aviation Psychology*, Dayton, OH. 32-37, CD-ROM.

Chiappe, D.L., Strybel, T.Z., Vu, K.-P.L. (2012). Mechanisms for the acquisition of situation awareness in situated agents. *Theoretical Issues in Ergonomic Science, 13, 625-647*.

Dao A. Q., Brandt, S., Battiste, V., Vu, K-P.L., Strybel, T.Z. & Johnson, W.W. (2009), The impact of automation assisted aircraft separation on situation awareness. In M.J. Smith and G. Salvendy (Eds.): *Human Interface, Part II*, HCII 2009 Lecture Notes in Computer Science 5618, 738–747.

D'Arcy, J-F. & DellaRocco, P.S. (2001). *Air Traffic Control Specialist Decision Making and Strategic Planning – A Field Survey*. New Jersey: Federal Aviation Administration.

Durso, F.T. & Dattel, A.R. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 137–154). Aldershot, UK: Ashgate.

Endsley, M. R. (1995a). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*, 32-64.

Endsley, M.R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors,* 37, 65-84.

Endsley, M.R., Sollenberger, R., & Stein, E. (2000). Situation awareness: A comparison of measures. *Proceedings of the Human Performance, Situation Awareness and Automation: User Centered Design for the New Millennium Conference*.

Gronlund, S.D., Ohrt, D., Dougherty, M., Perry, J.L. & Manning C.A. (1998). *Aircraft Importance and its Relevance to Situation Awareness.* Report DOT/FAA/AM-98/16. Federal Aviation Administration.

Salmon, P.M., Stanton, N.A., Walker, G.H. & Jenkins,D.P. (2009). *Distributed Situation Awareness: Theory, Measurement and Application to Teamwork*. Surrey, England:Ashgate Publishing.

Strybel, T. Z., Minakata, K., Nguyen, J., Pierce, R. & Vu, K-P., L., (2009), Optimizing online situation awareness probes in air traffic management tasks, In M. J. Smith and G. Salvendy (Eds.): Human Interface, Part II,

HCII 2009, *Lecture Notes in Computer Science*, 5618, 845-854.
Strybel, T.Z., Vu, K-P.L., Chiappe, D., Battiste, V., Johnson, W. & Dao, A-Q. (2011). *Metrics for situation awareness, workload, and performance: manual for online probe administration*. NASA/TM—2011–215991. National Aeronautics and Space Administration.
Strybel, T.Z., Vu, K.P. L., Battiste, V. & Johnson, W. (2013). Measuring the impact of NextGen operating concepts for separation assurance on pilot situation awareness and workload. *International Journal of Aviation Psychology, 23, 1-26.*

Acknowledgements

This project was supported by NASA cooperative agreement NNX09AU66A, Group 5 University Research Center: Center for Human Factors in Advanced Aeronautics Technologies (Brenda Collins, Technical Monitor).

Appendix A.

Sample probe stem questions with categories based on PCA and original task categories.

Sample Probe Stem Questions	Categories	Component
Estimate the lateral separation between the two closest co-altitude aircraft	Closest AC separation	Action Relevant
In the next 3 minutes (20 miles) will the majority of AC be coming from the West?	Global AC Traffic conditions	
In what general direction are the majority of overflights headed at this moment?		
How many conflicts will you resolve in the next 3 minutes (20 miles)?	General amount of upcoming ATC actions	
Will you solve any conflicts in the next 3 minutes (20 miles)?		
In what area will the next conflict occur if no further action is taken?	Next ATC action	
If AAL1320 is 2,000ft below its current altitude, how many AC will be in conflict with it in the next 3 minutes (20 miles)?		
How many miles before [callsign] reaches [waypoint]?	Lateral distance from AC to waypoint	Distance Relations
How many miles is [callsign] from [waypoint]?		
Is [callsign] higher in altitude than [callsign]?	Altitude distances between AC	
Is [callsign] lower in altitude than [callsign]?		
Will [callsign] be the next to cross [waypoint]?	Future lateral distances from AC to waypoint/AC	
Will [callsign] and [callsign] have less than 10 miles of separation if no further action is taken?		
From which direction will the next aircraft enter your sector?	AC Entering the sector	Low Priority
In the next 3 minutes (20 miles), will the majority of AC be entering the sector from the West?		
Have you moved any AC for traffic in the last 3 minutes (30 miles)?	Memory for prior conflicts	
How many conflicts have you resolved so far?		

MAN-MACHINE SYMBIOSIS IN AVIATION: NEW RISKS AND CAPABILITIES IN VIEW OF INFORMATION TECHNOLOGY EXPANSION

Oleksandr Petrenko
National Aviation University
Kyiv, Ukraine

The paper examines possibility of the impact of modern technology on the man-machine symbiosis in aviation. Presented some data concerning the perception of a pilots in a highly automated cockpit. The first results of empirical research make it possible to assume that aircraft systems can be customized to adjust to individual typological characteristics of operators. The paper discusses the language modality perspectives of a man-machine interaction. Creating modern sociotechnical systems aimed at a human at a completely new level has special requirements to an integration of designer and operational psychological outlooks.

Introduction

At all stages of aviation development there was a peculiar interaction between a man and an aircraft. S. Gerathewohl spoke about a sensitive display of psychological coupling of a man and an aircraft admitting that a piloting man experiences special mental conditions and feelings, including even an aesthetic component. The situation was like this in the middle of the 20th century. It was possible to talk about symbiosis characteristics at that time. Moreover, mentality reacts to new possibilities got by a man coupling his volitional and intellectual qualities with flying characteristics of a machine.

Sense of experiencing symbiotic coupling of a man with an aircraft which was called by Gerathewohl following his predecessors ‘flying feeling’ lies in including psychological mechanisms optimizing an interaction with a machine and an effective usage of psychological and physiological resources. Man’s subjective experiencing of coupling with technical systems is at the same time a process of man’s adaptation, his ‘adjustment’ to the interaction and a result of this adaptation.

An up-to-date stage of technology development gives a new wave to a man-machine symbiosis. There is a great distance between a Cessna pilot who is able to feel wings of his plane and a pilot working in highly automated cockpit of a modern airliner, who sometimes can experience a feeling of coupling his intellectual resources with resources of aircraft intellectual systems. Evidently, the harmony of this coupling can be experienced on the emotional level.

At an early stage of computer hardware development J.C.R. Licklider’s vision was that humans and machines could be coupled together and work interactively. He wrote: «Man-computer symbiosis is an expected development in cooperative interaction between men and electronic computers. It will involve very close coupling between the human and the electronic members of the partnership. The main aims are 1) to let computers facilitate formulative thinking as they now facilitate the solution of formulated problems, and 2) to enable men and computers to cooperate in making decisions and controlling complex situations without inflexible dependence on predetermined programs. In the anticipated symbiotic partnership, men will set the goals, formulate the hypotheses, determine the criteria, and perform the evaluations…

Preliminary analyses indicate that the symbiotic partnership will perform intellectual operations much more effectively than man alone can perform them». (Licklider, 1960, p. 4).

A movement to such a partnership in the era of information technologies expansion not only accelerates but become more controversial. It needs a comprehensive assessment of a sense of any ideas and constructive decisions related to the information technologies implementation to support and make a man-machine interaction more profound in aviation.

A Need for Innovation

So, the key manifestation of man-machine symbiosis is shown in achieving high efficiency by a man and reducing psychological costs of the work done by him.

It is known that a man in highly automated cockpit faces a lot of difficulties and a total psychological load on a man at certain periods of time can be even higher than in cockpits having lower level of automation.

The problem of man adaptation to activity conditions in highly automated cockpits can be solved by introducing an elaborate training system for men. Nevertheless, it is not completely tackled. We have conducted a survey (2013) questioning pilots of airlines based in Ukraine and flying Boeings 737-500. The pilots were offered to assess how modern level of cockpit automation has influenced their work using bipolar scales (Figure 1). General picture is positive taking into account average data. On the other hand, the grades of separate respondents proved that the problem is obvious (it is seen from a great difference in grades shown in the figure). It is important that according to the scale “more stressful – calmer work” some respondents put two marks at opposite sides despite given instructions.

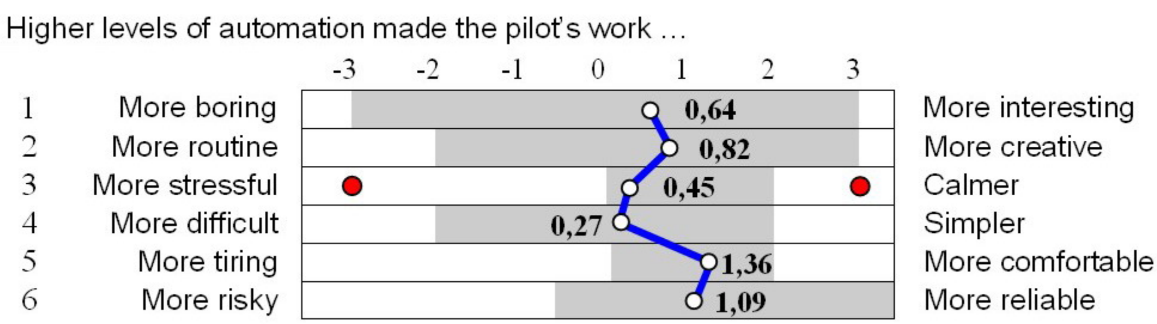


Figure 1. The perception of the pilots in a highly automated cockpit.

Giving answers to the question about difficulties occurring in the process of transition training for this type of an aircraft the respondents admitted they had some in the training process:

- Difficulties to understand logics of automation interaction (45% of respondents);
- Difficulties to transfer from multimember to two-member crew because of functions redistribution (12% of respondents).

Moreover, respondents claimed adaptation difficulties to a great number of control functions instead of executive; difficulties of visual information perceiving; doubts in automatics reliability.

The data we got prove that a pilots training system is not a comprehensive mean to solve the problem related to automation. It is obvious that the system is effective, but hardware and software development allowing better man-machine interaction is really needed.

The Development of Man-Machine Interfaces

The higher the level of aircraft cockpit automation the harder to assess risk of crew activity failure under the work conditions with patterned control procedures.

We are watching purposeful efforts of leading scientists of the world aimed at creating Cognitive Adaptive Man-Machine Interfaces (Dorneich and others).

Interfaces control is done taking into account instruments data about current man state in line with data about the condition of a controlled object, environment, situation which allows to avoid man's information overload and harmonize his interaction with the machine.

Our studies are aimed at exploring possibilities of man-machine interaction 'adjustment' to typological characteristics of an operator.

To solve the problem the regularities of different psychic functions interference must be studied and formalized as well as individual and typological differences of such interference. Finding out these regularities will allow us not only give a current assessment of crew reliability but realize adaptive information models aimed at minimizing risks of faulty actions connected with informational overload of a man.

To conduct our survey we use integrated stand giving an opportunity to simulate joint work of two operators. Each of them is engaged simultaneously in different contours of control and processes information of different modality performing combined tasks of different goals under the conditions of current language interaction between partners (Figure 2).

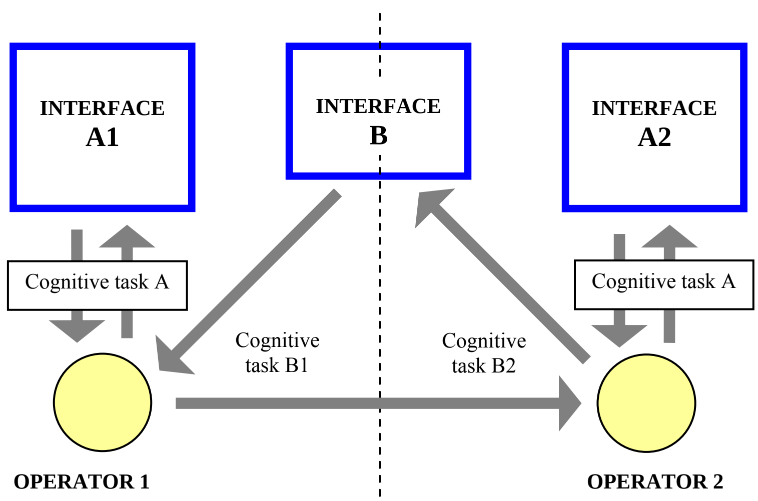


Figure 2. A scheme of integrated stand to study interference of different psychic functions of an operator.

Judging by the first results of the begun empiric studies we can suggest that typological characteristics of a man-operator are identified by a wide range of individual peculiarities including individual cognitive style, psychic asymmetry profile, neurodynamic characteristics and some personality features.

Information about typological characteristics of crew members can be used together with the information about a current condition. While a current state is subject to instrumental monitoring, data about typological characteristics can be entered into airborne systems before the beginning of work with the help of personal key together with the information kept in it.

Transformation of Man Tasks and Requirements to His Features

Creation of adaptive man-machine interfaces allowing to reduce the risk of informational overload of a man is one of the tasks to develop man-machine systems.

In many cases when human factor is a cause of an emergency we can admit a lack of quality of man heuristic activity (scenario forecasting, understanding of a flying situation in general, working out new patterns of behavior, making decisions feeling lack of information). This is an activity which man performs better than a machine but airborne systems can assist a man. It goes not only about unloading a man giving him more opportunities to solve heuristic tasks but providing him with a direct help in the process of their tackling. Achieving such level of interaction between a man and a machine we can speak about a brand new situation when a machine becomes a symbiotic partner of a man and perceives him as another crew member or as an extension of his own mind.

In these cases we deal with special teams which are called hybrid. It was shown that working successfully in highly automated Human-Machine-Interfaces (HMI), i.e. in a "hybrid team", demands different aspects of personality and attitudes than working with a human partner (Eschen-Léguedé, Knappe, Keye, 2011).

Language Modality Perspectives of a Man-Machine Interaction

Solving heuristic problems in the cockpit a human usually deals with more generalized levels of reality than levels represented by instrument board indicators. A question arises which means of information transmission can be used more effectively in such situations.

We believe verbal messages are adequate to these levels as speech as means of information transmission is notable for peculiar possibilities of generalizing and informational capacity. That's why the most important element of systems providing help to a human in overcoming flight difficulties can be subsystems of realizing verbal man-machine interaction.

The analysis of possibilities capable to provide such systems and their implementation in aviation and ways of reducing risks was conducted when a level of IT and computers development was far from the up-to-date one (Ushakova, Pavlova, Zachesova, 1989). At that time the creation of hybrid teams was out of question and the interest to interaction of communication channel and technics was connected to systems of speech warning and voice control. Nevertheless, the main defined provisions are still urgent which make a good basis for continuing research in the direction.

Unlike a talk exchange between people which being natural for them because of their nature is used to solve any including the simplest task, a talk exchange between a human and a technical system is mostly beneficial while solving difficult tasks to assess situations and make decisions. To realize such talk exchange it is necessary to use special phraseology expressed in laconic word forms suitable to transfer reliably highly generalized messages concerning different classes of flight situations but at the same time unambiguous in every possible context. There is a need to prevent faulty language commands of an operator and misunderstandings which are likely to occur without taking specific measures.

We questioned pilots flying up-to-date aircraft about their attitude to a possibility of talk exchange with onboard systems. Expecting to get mostly skeptical attitude of pilots to such an opportunity, in fact only 27% of respondents expressed skeptical views. At the same time pilots treated more positively the perspective of developing systems of voice communication of a machine to a human than an introduction of voice systems to control machines. Another regularity expresses positive attitude to a possibility of voice communication with onboard systems which positively correlated with positive attitude to cockpit automation.

Suggesting that the attitude to voice communication systems is stipulated by individual peculiarities of cognitive sphere we compared groups of pilots who treated such systems mostly positively and mostly negatively taking into account individual cognitive style data, verbal and logical thinking and creative abilities. For this purpose we used a cognitive style questionnaire (A.Harrison, R.Bremson, adaptation by A.Alekseev) allowing to assess a level of an individual inclination to synthetic, idealistic, pragmatic, analytical and realistic styles as well as verbal and logical test and verbal creativity test. It turned out that pilot groups who positively and negatively treat the perspective of wide usage of communication systems in man-machine interfaces drastically differ by inclination to analytical cognitive style data (pilots having more positive attitude to using voice communication systems demonstrated higher results of analytical style). There were no differences found in other styles as well as in verbal-logical thinking and verbal creativity.

Moreover, it was found out that pilots giving more positive grades as to the introduction of voice communication systems believe that the work in automated cockpits is more interesting (drawing 1, scale 1) comparing with the work in cockpits of previous generation aircraft. This fact and found correlation of pilot treatment of automation and language man-machine interaction can be interpreted as a proof that the idea of interactive communication systems design in man-machine interfaces correlates with pilots understanding of general trends of cockpit automation and doesn't oppose to their professional psychological outlook.

The Ideological Aspects of Innovation

Introducing automation systems enables automatics developers to be indirectly present in the cockpit together with the crew.(Golikov, Kostin, 1996). Creating modern sociotechnical systems aimed at a human at a completely new level has special requirements to an integration of designer and operational psychological outlooks. A problem of these outlooks correlation has an objective background connected with different professions owing their perceptional space (Strelkov, 2001).

Technics designers have covered a long way from a machinecentric approach to an anthropocentric one. Nevertheless, it is too early to talk about problem solving. However, designer's activity pushes a developer if not to a machinecentric approach then to a simplified understanding of anthropocentrism in the psychology framework of cognitive processes.

Professional stereotypes encourage aircraft developers to solve problems caused by automation, using further actions of automation which in its turn leads to further model transformation of cockpit crew actions. Such transformation causes specific difficulties for experienced crews adding to the difficulties characteristic of human activity in highly automated cockpit.

As an example of a field where aircraft developers and pilots positions are coordinated we can take a flight-test unit of aircraft manufacturing plant. So, at Antonov designing bureau we can find examples of fighter pilots taking an active part in position transformation of aircraft developers towards approaching the outlook typical for cockpit crew. To make the matter successful communication between developers and producers must take place at the stage of designing.

In general we claim that new risks and new opportunities connected with further cockpit automation are integral. Ultimately, the task is to prevent possibilities transformation into risks but to encourage risks transformation into possibilities. For this purpose it is necessary not to lose a personality of a professional as a system-forming phenomenon behind other narrow-specialized issues. Professional outlook, realizing his role and his limits, his heuristic potential are of same importance under the conditions of 'hybridization' of social and technical systems as improving of man-machine interfaces and activity algorithms.

Acknowledgements

Any opinions, recommendations, findings, or conclusions expressed herein are those of the authors and do not necessarily reflect the views of the National Aviation University.

References

Dorneich, M.C., Passinger, B., Beekhuyzen, M., Hamblin, C., Keinrath, C., Whitlow, S., & Vašek, J. (2011). The Crew Workload Manager: An Open-loop Adaptive System Design for Next Generation Flight Decks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* Las Vegas, NV, September 19-23 (pp. 16-20).

Eschen-Léguedé, S., Knappe, K., Keye, D. (2011). Aspects of personality in highly automated Human-Maschine-Teams - Development of a questionaire. In: *Reflexionen und Visionen der Mensch-Maschine-Interaktion* - Aus der Vergangenheit lernen, Zukunft gestalten Reihe 22. Mensch-Maschine-Systeme, 33. ZMMS (pp. 459-464).

Gerathewohl, S. (1956). Psychology of man in an aircraft. Moscow, IL. (pp. 181-193).

Golikov, Yu.Ya, Kostin, A.N. (1996). Psychology of control technics automation. – Moscow, IP RAS (160 p.).

Harrison, A.F., Brainson, R.M. (1984) The art of thinking. N. Y., Berkley Books (pp. 189-193).

Licklider, J. C.R. (1960). Man-Computer Symbiosis. In: *IRE Transactions on Human Factors in Electronics*, Vol. HFE-1 (pp. 4-11).

Strelkov, U.K. (2001). Engineering and professional psychology. Moscow, Akademia, (360 p.)

Ushakova, T.N., Pavlova, N.D., Zachesova, I.A. (1989). Human speech in communication. Moscow, Nauka (pp. 172-183).

ADAPTIVE CONTROLLER ADAPTATION TIME AND AVAILABLE CONTROL AUTHORITY EFFECTS ON PILOTING

Anna Trujillo and Dr. Irene Gregory
NASA Langley Research Center
Hampton, VA

Adaptive control is considered for highly uncertain, and potentially unpredictable, flight dynamics characteristic of adverse conditions. This experiment looked at how adaptive controller adaptation time to recover nominal aircraft dynamics affects pilots and how pilots want information about available control authority transmitted. Results indicate that an adaptive controller that takes three seconds to adapt helped pilots when looking at lateral and longitudinal errors. The controllability ratings improved with the adaptive controller, again the most for the three seconds adaptation time while workload decreased with the adaptive controller. The effects of the displays showing the percentage amount of available safe flight envelope used in the maneuver were dominated by the adaptation time. With the displays, the altitude error increased, controllability slightly decreased, and mental demand increased. Therefore, the displays did require some of the subjects' resources but these negatives may be outweighed by pilots having more situation awareness of their aircraft.

Adaptive control in flight applications has a long and rich history dating back to the 1950s. Currently, adaptive control is beneficial for highly uncertain, and potentially unpredictable, flight dynamics characteristic of upset recovery or damage induced on transport as well as high-performance aircraft. Some of the recent flight experiences of pilot-in-the-loop with an adaptive controller have exhibited unpredicted interactions (Bosworth & Williams-Hayes, 2007; Page, Meloney, & Monaco, 2006). In retrospect, this is not surprising once it is realized that there are now two adaptive controllers interacting, the traditional software adaptive control system and the pilot. The pilot is another entity that may affect the attitude of the vehicle (definition of a control system), and the pilot's method of controlling may change due to slowly varying or uncertain system parameters. One hypothesized reason for the pilot-in-the-loop with an adaptive controller interactions is that it is due to the pilot not realizing what the adaptive controller is doing and what the limits of the adaptive controller are.

The experiment objectives were to determine (1) how the adaptation time of the controller affects pilots and (2) how pilots want information about the control authority (or maneuver capability) available to them transmitted.

Method

This experiment looked at whether an adaptive controller helps pilots during control surface failures and whether displays indicating how close the vehicle is to reaching the limit of safe maneuver envelope were helpful before, during, and after the control surface failures. The limits indicated to the subjects were bank angle, vertical velocity, and aircraft speed (Trujillo & Gregory, 2013; Wilborn, 2001). Furthermore, these variables were used in two displays designed to inform the pilot of available maneuverability envelope. These displays were then used in a human-in-the-loop experiment to look at their effects on pilot performance with aircraft surface failures during cruise phase while initiating a climb, descent, or a heading change maneuver. These maneuvers were indicated on the primary flight display (PFD) via the flight director.

Simulation Environment

The physical setup of the simulator incorporated an out-the-window view in the upper center 30-inch diagonal screen and four 20-inch touchscreens below the out-the-widow screen. The middle-left touchscreen depicted the PFD and the middle-right touchscreen depicted the engine indication display (EID). The far-left touchscreen contained the control authority display when present and the far-right touchscreen displayed the after run questions. Subjects flew the aircraft with a right-handed joystick.

Independent Variables

Display Type. The two displays tested were the dial display (Figure 1) and the circle display (Figure 2). In both displays, the information shown was the same but the format was different. In each display, a green wedge filled in from zero the percentage of available safe maneuver envelope used in the task. For example, for vertical velocity (VVel) in Figure 1, the aircraft is descending at 100% or more of available 3000 ft/min. When the available control authority changed from normal due to failure, the displayed number went from white to cyan in color and the limit value changed to the newly available one. For example, for minimum speed (Min Spd) in Figure 2, the aircraft's safe minimum speed is now 120 kts as indicated by the cyan number.

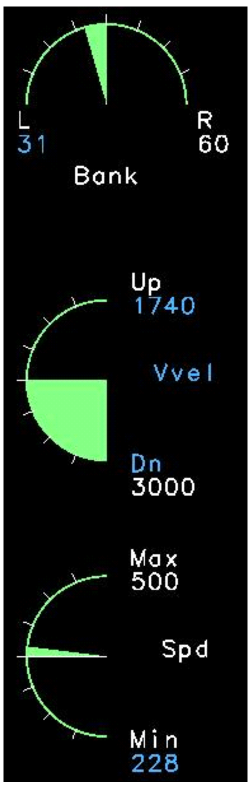


Figure 1. Dial Display

Adaptation Time. Each subject experienced four adaptation times: zero seconds, three seconds, seven seconds, and no adaptation (Never). These times indicated how long it took the adaptive controller to settle to new aircraft dynamics and are based on the response of aircraft dynamics. Zero seconds indicated the fastest possible adaptation time, essentially the processor speed. Three seconds was chosen because with this time, the subject might notice the controller adapting. As for seven seconds, this was chosen because the subject should notice the controller adapting.

Subjects. The seventeen subjects were an average of 48±10 years old with the youngest 29 years old and the oldest 61 years old. All of them were airline transport rated pilots with an average of 26±11 years of flight experience (minimum flight experience = 7 years and maximum flight experience = 45 years) and an average of 10,706±7164 hours of flight experience (minimum flight hours = 2,100 and maximum flight hours = 23,400).

Dependent Variables

The primary dependent variables involved flight technical data. In particular in the lateral axis was cross track error, the difference between current aircraft position and commanded position, and roll error, the difference between current bank angle and commanded bank angle. In the longitudinal axis was altitude error, the difference between current aircraft altitude and commanded altitude, and pitch error, the difference between current aircraft pitch angle and commanded pitch angle.

Two other secondary dependent variables involved subjective ratings by the participant. After each run, subjects provided a Cooper-Harper (CH) handling qualities (HQ) rating (Cooper & Harper, 1969; Harper & Cooper, 1986; Trujillo, 2009). After certain runs, subjects also gave a NASA-TLX workload rating (Hart & Staveland, 1988; Trujillo, 2011).

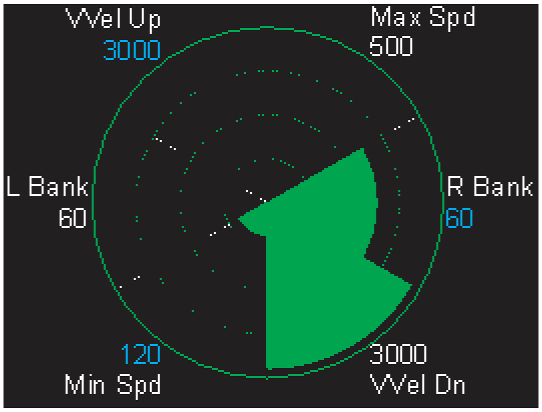


Figure 2. Circle Display

Procedure

Each subject had several runs without the new displays (None), then several runs with either the circle or dial display, and finally several runs with the other display. During each run, flight technical data were recorded. After each run, subjects gave a CH rating and a NASA-TLX workload rating. After all the data runs were completed, subjects filled out a final questionnaire asking them about their preferences on the information in the displays and displays themselves.

Results

Flight Technical Data

Lateral Error. Adaptation time was significant for cross track error during ($F_{(2,868)}$=36.90, p≤0.01) and immediately after ($F_{(2,873)}$=28.36, p≤0.01) the control surface failure and roll error was significant during ($F_{(2,868)}$=26.07, p≤0.01), immediately after ($F_{(2,873)}$=3.79, p≤0.03), and after ($F_{(3,1157)}$=7.54, p≤0.01) the control surface error. In general, the 3 sec adaptation time was associated with the least cross track error (Figure 3) and roll error (Figure 4). This may indicate that while subjects were able to follow the flight path with no adaptation, fine motion was compromised.

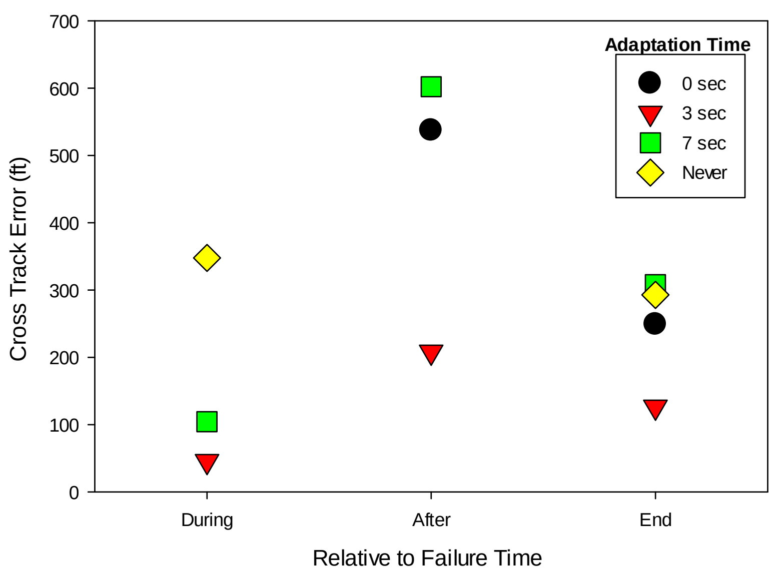


Figure 3. Cross Track Error During, Immediately After, and at the End by Adaptation Time

Unsurprisingly, when the adaptive controller never engaged, subjects improved their performance for both cross track error and roll error as time progressed. This suggests that indeed subjects were adapting to the vehicle's new dynamics. Also note that when the adaptive controller never engaged, the lateral errors were greater than with an adapting controller. In fact, the roll error with the adaptive controller was less than the roll error before the control surface failure. This indicates that the adaptive controller was helping the subjects control the aircraft and given enough time, the cross track error decreased.

Longitudinal Error. Adaptation time was significant for altitude error during ($F_{(2,868)}$=201.25, p≤0.01), immediately after ($F_{(2,873)}$=47.83, p≤0.01), and at the end of the run ($F_{(3,1157)}$=41.83, p≤0.01) and was significant for pitch immediately after ($F_{(2,873)}$=10.92, p≤0.01) and at the end ($F_{(3,1157)}$=12.73, p≤0.01) of the run. The zero and seven second adaptation times were associated with the least altitude error (Figure 5) and pitch error (Figure 6).

Again, as with lateral error, when the adaptive controller never adapted, subjects improved their

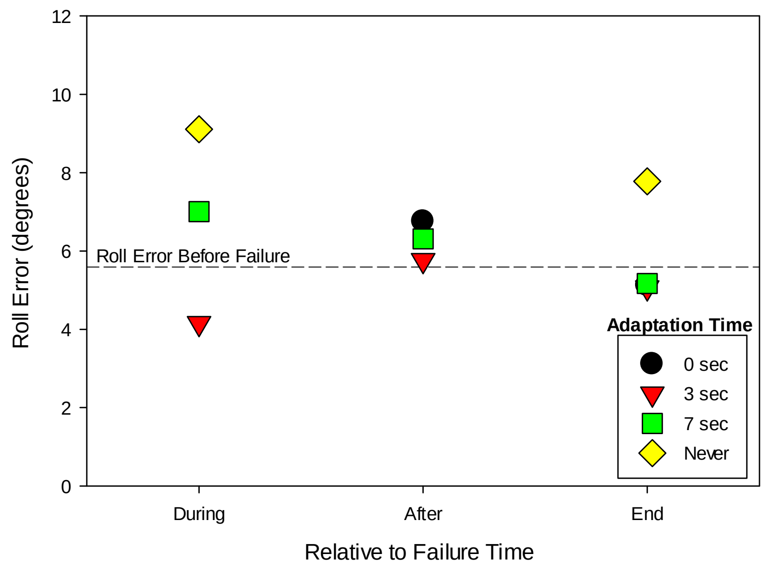


Figure 4. Roll Error During, Immediately After, and at the End by Adaptation Time

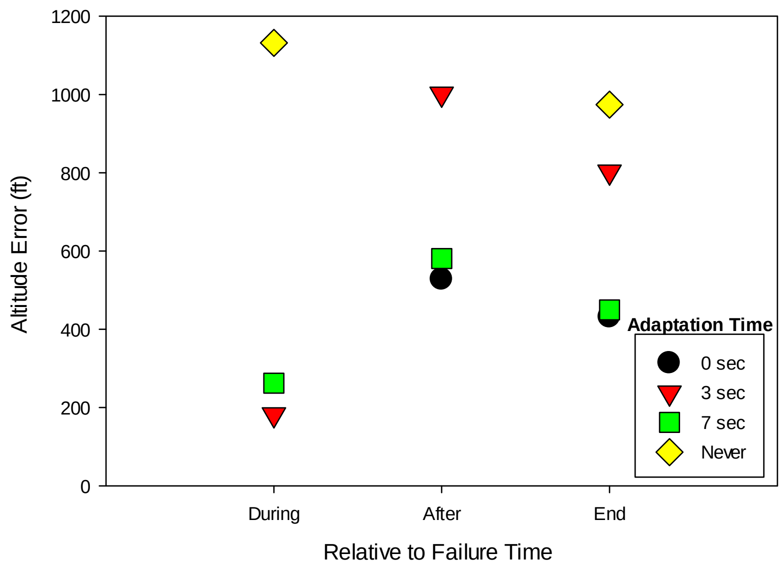


Figure 5. Altitude Error During, Immediately After, and at the End by Adaptation Time

performance for both altitude error and pitch error as time progressed. As before, when the adaptive controller never adapted, the longitudinal errors were greater than with an adapting controller.

Display type was significant for altitude error before the failure ($F_{(2,1175)}$=4.05, p≤0.02). As can be seen in Table 1, the least amount of altitude error is associated with no display. Although not significant, this trend also held for altitude error immediately after the failure and at the end. Without the display present, subjects were able to perhaps expend more attention on the PFD maintaining aircraft path.

Table 1.
Altitude Error Before the Failure by Display Type

Display Type	**Altitude Error (ft)**
None	42.98
Circle	56.80
Dial	59.77

Cooper-Harper Handling Qualities Ratings

Both display type (Figure 7) and adaptation time (Figure 8) were significant for the CH rating (display: $F_{(2,1157)}$=4.06, p≤0.02 ; $F_{(3,1157)}$=17.56, p≤0.01). With the new displays, the CH ratings increased slightly indicating poorer handling qualities. This decrease in HQ may be due to the subjects having to expend resources to process the new displays rather than focusing on maintaining aircraft control. The CH ratings improved with the adaptive controller. As seen with the flight technical data, the three second adaptation time appears to improve the CH ratings the most.

Workload

Adaptation time was significant for workload ($F_{(3,764)}$=5.29, p≤0.01). As can be seen in Table 2, no adaptation had a higher workload than the other three adaptation times. This suggests that the adaptive controller did decrease the workload of the subject during control surface failures.

As for display type, it was only significant for mental workload ($F_{(2,764)}$=3.51, p≤0.02). The two

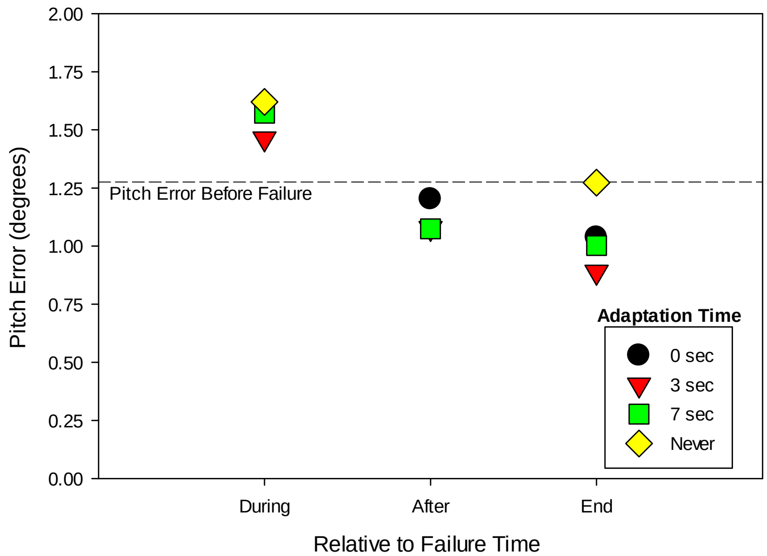


Figure 6. Pitch Error During, Immediately After, and at the End by Adaptation Time

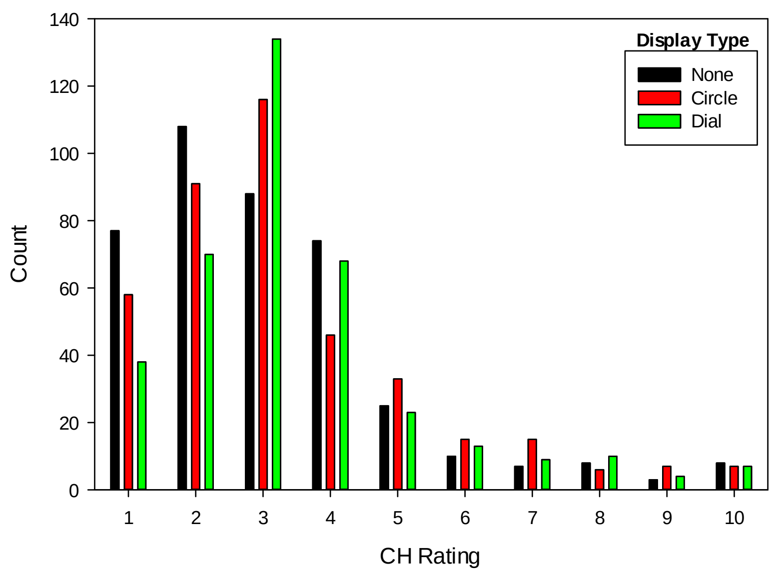


Figure 7. Frequency of CH Rating by Display Type

Table 2.
Select NASA-TLX Ratings by Adaptation Time and Display Type

Adaptation Time	**Workload**
0 sec	21.64
3 sec	24.42
7 sec	22.52
Never	35.29
Display Type	**Mental Demand**
None	23.32
Circle	29.04
Dial	25.99

Note. 0 = Low; 100 = high ratings.

displays did increase mental demand (Table 2) indicating that the displays did require some mental resources from the subjects. Of the two displays, the dial display required less of an increase in mental resources. This was most likely because the dial-type display was a familiar format to the subjects whereas the circle display was new to them and not used in current flight decks.

Conclusions

Adaptive control is beneficial for highly uncertain, and potentially unpredictable, flight dynamics that are characteristic of upset recovery or damage induced on transport or high-performance aircraft. But with an adaptive controller and a pilot, there are now two adaptive systems interacting, the traditional software adaptive control system and the pilot. The pilot controls the attitude of the vehicle (definition of a control system) and the method of control may change due to slowly varying or uncertain system parameters. This experiment looked at whether an adaptive controller helps pilots during control surface failures and whether displays indicating how close the vehicle is to reaching the limit of safe maneuver control authority were helpful before, during, and after the control surface failures. The limits indicated to the subjects were bank angle, vertical velocity, and aircraft speed.

Results indicate that an adaptive controller that takes three seconds to adapt helped the pilot most when looking at lateral and longitudinal errors. This adaptation time may be short enough to cause minimal interactions with the pilot possibly adapting but long enough to have the pilot realize that the aircraft has a control problem. Another possibility is the instantaneously adapting controller was too

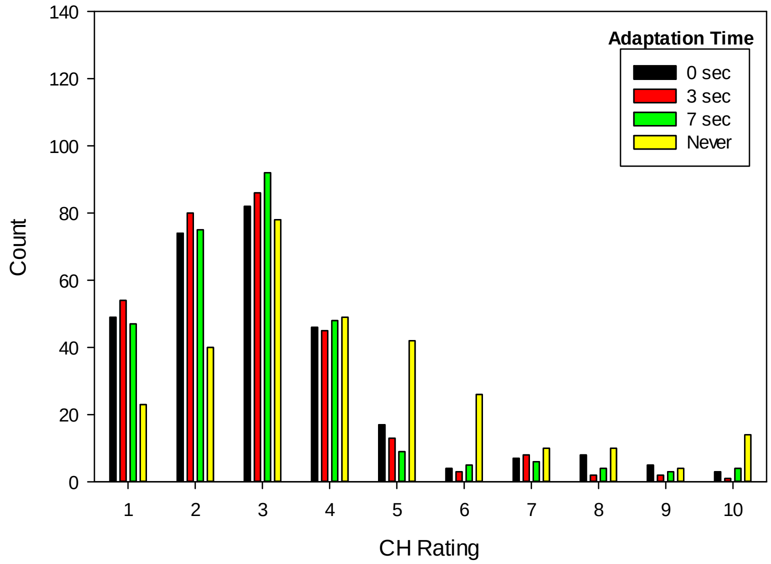


Figure 8. Frequency of CH Rating by Adaptation Time

sensitive for the pilots to fly comfortably. Handling quality ratings improved with the adaptive controller, again for the three second adaptation time, while workload decreased with the adaptive controller. The adaptive controller was helping the subjects control the aircraft and given enough time, subjects' lateral and longitudinal errors continued to decrease to be on par with errors before any control surface failures occurred. Even without an adaptive controller, subjects improved their performance by adapting to the vehicle's new dynamics.

The effects of the displays showing the percentage of available safe maneuver envelope used in the task were dominated by the adaptation time. With the displays, altitude error did increase along with a slight decrease in HQ. The additional display also required increased mental demand. Therefore, the displays did require some of the subjects' resources but these negative effects are minimal and may be outweighed by pilots having more situation awareness of a control problem with the aircraft and the failure's effects on their ability to control the aircraft.

Acknowledgements

The authors would like to recognize Mr. Lucas Hempley of Northrop Grumman Technical Services. Without his help and patience, this experiment would have never run.

References

Bosworth, J. T., & Williams-Hayes, P. S. (2007). *Flight Test Results from the NF-15B Intelligent Flight Control System (IFCS) Project with Adaptation to a Simulated Stabilator Failure*. Paper presented at the AIAA Infotech@Aerospace 2007 Conference and Exhibit, Rohnert Park, CA.

Cooper, G. E., & Harper, R. P. (1969). The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities (pp. 52): AGARD.

Harper, R. P., & Cooper, G. E. (1986). Handling Qualities and Pilot Evaluation (Wright Brothers Lecture in Aeronautics). *Journal of Guidance, Control, and Dynamics, 9*(6), 515-529.

Hart, S. G., & Staveland, L. E. (1988). Development of a NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. S. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: Elsevier Science Publishers B. V.

Page, A. B., Meloney, E. D., & Monaco, J. F. (2006). *Flight Testing of a Retrofit Reconfigurable Control Law Architecture Using an F/A-18C*. Paper presented at the AIAA Guidance, Navigation, and Control Conference and Exhibit, Keystone, CO.

Trujillo, A. C. (2009). *Paper to Electronic Questionnaires: Effects on Structured Questionnaire Forms*. Paper presented at the HCI International 2009, San Diego, CA.

Trujillo, A. C. (2011). Evaluation of Electronic Formats of the NASA Task Load Index (pp. 33). Hampton, VA: NASA Langley Research Center.

Trujillo, A. C., & Gregory, I. M. (2013). *Pilot Preferences on Displayed Aircraft Control Variables*. Paper to be presented at the HCI International 2013, Las Vegas, NV.

Wilborn, J. E. (2001). An Analysis of Commercial Transport Aircraft Loss-of-Control Accidents and Intervention Stategies (pp. 23).

SOUTH AFRICAN AIRLINE PILOTS' PERCEPTIONS OF ADVANCED FLIGHT DECK AUTOMATION

Dr. Preven Naidoo.

Dept. of Human Resource Management

University of Pretoria, South Africa

This article reports on the construction of the Automation Attitude Questionnaire (AAQ), to assess airline pilots' perceptions about operating advanced automated aircraft. A total of 262 airline pilots from a large South African carrier participated in the validation of the questionnaire. A five-factor measurement model was established by using exploratory factor analysis. The five factors associated with perceptions of advanced automated systems were labelled as: Comprehension, Training, Trust, Workload, and Design. The Cronbach's alpha coefficients and the mean inter-item correlation of each factor were highly satisfactory and confirmed the homogeneity and unidimensionality of the five-factor solution for the AAQ.

Advancements in computer and processing technology compounded by economic demands have given rise to increasingly electronic flight deck design on aircraft. During the 1980s there was a rapid development of automated apparatus, which were incorporated, in large aircraft. Significant developments included inertial reference systems (IRS), flight guidance systems, auto-throttle/thrust systems, data and flight management systems (FMS), and various crew alerting systems (Boeing EICAS and Airbus ECAM) (Risukhin, 2001; Wiener, 1989). In general terms, the new technology manifested itself as 'the glass cockpit' '(displays driven by computer graphic systems)' (Wiener, 1988, p. 435). The integration of automation into commercial aircraft flight decks has contributed to greater efficiency, productivity and overall safety (Wiener, 1993).

The introduction of automation to the modern aircraft flight deck, however, has also resulted in a debate 'for' or 'against' such automation (Risukhin, 2001). Wiener (1989, p.1) indicated that as the level of automation increased there was 'a growing discomfort that the cockpit may be coming too automated.' Human factor issues such as poor interface design, pilot complacency and over-reliance on automation, deteriorating flying skills and diminished situational awareness began to be considered (Billings, 1997; Palmer, 1995; Parasuraman, & Riley, 1997; Wood, 2004).

The results of the survey related to airline pilots' perceptions towards operating advanced automated flight deck systems is presented as a two part article series. In part one the need for the study and the construction of the Automation Attitude Questionnaire (AAQ) is described. In the second article the relationship between the airline pilots' perceptions of flight deck automation and their personal characteristics and operational position are explored.

In the present study the researchers seek to identify the core human factor issues related to advanced flight deck automation and to construct a valid and reliable instrument to measure the perceptions of airline pilots towards operating advanced automated aircraft.

Background

Humans operating flying machines started only a century ago when Orville Wright famously took the controls of the magnificent Wright 'Flyer' in 1903 and became airborne for a distance approximately the length of a modern Boeing 747-400 (Crouch, 2003). However, the wooden flyer had initially stalled during the Wright brothers' first attempt at flight, and this was arguably the first ever man-machine aviation-related accident, as described in this paragraph by Wilbur Wright after the incident: '…the power is ample, and but for a trifling error due to lack of experience with this machine and this method of starting, the machine would undoubtedly have flown beautifully' (Crouch, 2003, p. 43). The aircraft that Orville flew on that important day in history was also the starting point in the aviation advances from an extremely manual mode of flight to the point of automation and computerisation found on the modern flight deck today.

A vast difference in design exists between the aircraft being built today and the wooden device flown by the Wright brothers over a century ago. The present advanced flight deck incorporates flight data information on cathode ray tubes (CRTs) and liquid crystal displays (LCDs) – the main reason that many observers refer to these systems as 'glass cockpits' (Risukhin, 2001). According to Risukhin (2001), the complete digitised flight deck system consists of electronic attitude director indicators (EADIs); electronic horizontal situation indicators (EHSIs); data management systems (FMS) and symbol generators to drive the electronic indicators; navigation system control and display units (ND); and air data systems. Various crew alerting systems (in Boeing the EICAS, and in Airbus the ECAM) are incorporated on the modern flight deck to support pilots to operate aircraft more safely in today's congested airspace. This includes for example a Traffic Collision and Avoidance System (TCAS) and Controlled Flight into Terrain (CFIT) avoidance equipment, such as the Enhanced Ground Proximity Warning System (EGPWS) technology

The modern jet airliner has evolved from that first flight of the imagination of Leonardo Da Vinci and is the culmination of that dream, integrating in its complex technology almost all of humankind's scientific thought to date (Paolo, 2001). Only computer technology has paralleled aviation in its rapid evolutionary advance; and the modern jet airliner is the proud heir of both fields of human endeavour, combining aviation and computer technology in an elegant fusion. The quantum leap in flight deck design and layout is evidenced from the impressive glass panel displays found on the modern advanced aircraft.

Although flight deck automation has been well received by the aviation industry and pilots, many human factor issues have been raised. Research suggests that the increased presence of computers (such as flight management computers, FMC's) on board modern flight decks have resulted in some flight crew members' spending an increasing amount of 'heads-down' time during critical phases of flight, a key contribution to distractions resulting in an incident or accident (Damos, John & Lyall, 2005). Traditionally, the operation of analogue flight deck aircraft meant that pilots were often making an exceedingly large number of minute mistakes. The modern advanced flight deck incorporates highly sophisticated computers which now take care of the mundane or routine aircraft operations. Any mistakes committed by the human operator on a modern flight deck are more likely to result in a catastrophic disaster (Edwards, 1988). For example, the use of reduced or flexible thrust take-offs have become an everyday method of ameliorating wear and tear on jet engines. However, an error in the input of the correct temperature (a two-digit number) into the flight management computer may result in disaster if the aircraft fails to accelerate during the take-off phase (when the assumed temperature is far higher then is actually required). This highlights the fallibility of the basic computer-human dyadic. Experts in the field refer to this as GIGO or 'garbage-in-garbage-out' (Damos et al., 2005). In other words, this dyadic is only as strong as the weakest link, the human being. There is ample proof that human error is a root cause of accidents in complex systems (NTSB, 2009).

Analyses of the reasons and variables implicated in aircraft incidents and accidents indicated that the rise in the number of aircraft accidents in the past 20 years have emerged as a symptom of the increasing use of automation throughout the world aviation industry. (FAA, 1996; Ishibashi, Kanda & Ishida,1999; Skitka, Mosier, Burdick & Rosenblatt,2000). Examples of accidents were the breakdown in flightcrew/automation coordination were the main contributor factor included an Airbus A300-600 operated by China Airlines that crashed at Nagoya in 1994; a Boeing 757 operated by American Airlines that crashed near Cali, Columbia in1995; and recently a Boeing 737-800 operated by Turkish Airlines that went down in a muddy field less than a mile short of the runway at Amsterdam's Schiphol airport shortly before it was due to land on 25 February 2009.

However, it is a statistically documented fact that commercial jet air travel is still the safest mode of transportation known to mankind (NTSB, 2009) – and this is why the public is often confused, shocked and horrified by any accidents involving advanced automated aircraft (Risukhin, 2001). In trying to answer the question of why such accidents happen, one must critically analyse the new problems and challenges presented by the introduction and utilisation of computerisation in aircraft (Risukhin, 2001). It is therefore important to understand the human perception of an advanced automated environment in the execution of the safe operation of these aircraft. A negative or false perception of technology and automation may have an adverse impact on safety issues, but over-familiarity with the system may bring about boredom, fatigue and complacency.

Human factor issues and automation

The term "automation" has for some time been difficult to define, although many researchers working in the field have agreed that the term "…generally means replacing human functioning with machine functioning", whilst in the term flight deck automation "…we generally mean that some tasks or portions of tasks performed by the human crew can be assigned, by the choice of the crew, to machinery" (Wiener, 1989, p. 121). Funk, Lyall, Wilson, Vint, Niemczyk, Surotegu and Owen (1999, p. 56) also indicated that "Automation is the allocation of functions to machines that would otherwise be allocated to humans".

The integration of automation into commercial aircraft flight decks has contributed to economic profitability of airlines, has allowed more efficient flight path management and reduced the number of crew needed for operation from three to two simultaneously improving overall safety (Kabbani 1995; Rudisill, 1995; Wiener, 1993). However, while there were benefits from automation, several unique human factor issues have been raised by both the civil aviation authorities and by human behaviour experts (Billings, 1997; FAA, 1996; Kabay, 1996; Palmer, 1995; Parasuraman, & Riley, 1997, Wood, 2004). These human factor issues relate to concerns about poor interface design, pilot complacency and over-reliance on automation, a loss of manual flying skills, and pilots' lack of understanding of the new equipment, mode errors and 'automation surprises'.

Aviation human factors literature (FAA, 1996; Funk, Lyall, & Riley, 1995; Funk and Lyall, 1999, 2000; Funk et al., 1999; Kabbani, 1995, Rudisill, 1995; Sarter, 1995, 1996; Sherman, Helmreich & Merritt, 1997; Wood, 2004) refers to many problems and concerns with flight deck automation. The most comprehensive study on automation issues affecting pilots operating advanced automated aircraft was coordinated and reported by Funk, Lyall, & Riley, 1995, Funk and Lyall, 1999, 2000, and Funk et al., 1999. The contribution of various aviation scholars and aviation operators resulted in research output that have been intergraded and documented in the public domain and can easily be accessed via the worl wide web. Various contributors have identified 92 critical issues affecting pilots with regard to automation and the operation of advanced aircraft. The 92 issues which were identified by analysing different sources and surveying actual operators are too cumbersome to be listed in its entirety for this article. However, ranked by the sum of evidence strengths, the following 10 automation issues were acknowledged by Funk and Lyall (2000:5) as the most significant factors affecting pilots operating advanced aircraft.

- *Understanding*: ‘Pilots may not understand the structure and function of automation or the interaction of automation devices well enough to safely perform their duties’.
- *Mode awareness*: ‘The behavior of automation devices -- what they are doing now and what they will do in the future based upon pilot input or other factors -- may not be apparent to pilots, possibly resulting in reduced pilot awareness of automation behavior and goals’.
- *Complacency/Trust*: ‘Pilots may become complacent because they are overconfident in and uncritical of automation, and fail to exercise appropriate vigilance, sometimes to the extent of abdicating responsibility to it. This can lead to unsafe conditions’.
- *Design*: ‘Displays (including aural warnings and other auditory displays), display formats, and display elements may not be designed for detectability, discriminability, and interpretability. This may cause important information to be missed or misinterpreted’.
- *Training*: ‘Training philosophy, objectives, methods, materials, or equipment may be inadequate to properly train pilots for safe and effective automated aircraft operation’.
- *Inappropriate usage*: ‘Pilots may use automation in situations where it should not be used’.
- *Complexity*: ‘Automation may be too complex, in that it may consist of many interrelated components and may operate under many different modes. This makes automation difficult for pilots to understand and use safely’.

- *Surprise events*: ‘Automation may perform in ways that are unintended, unexpected, and perhaps unexplainable by pilots, possibly creating confusion, increasing pilot workload to compensate, and sometimes leading to unsafe conditions’.
- *Dissemination of information*: ‘Important information that could be displayed by automation is not displayed, thereby limiting the ability of pilots to make safe decisions and actions’.
- *Reduced skill*: ‘Pilots may lose psychomotor and cognitive skills required for flying manually, or for flying non-automated aircraft, due to extensive use of automation’.

A qualitative analysis of the freehand comments of 400 pilots by Rudisill (1995), based on the results from an international survey conducted by James, McClumpha, Green, Wilson and Belyavin (1991a) revealed issues closely aligned with the top ten issues as reported in the automation data base. The results of this examination resulted in the following issues and recommendations from Rudisill (1995):

- *General issues related to automation*: The general agreement among participants was that automation was a good thing. However, concerns were raised that inexperienced pilots may be led into a false sense of security by the automatics. One solution proposed to resolve this issue is to provide mechanisms for inexperienced pilots to gain and develop a firm base in piloting skill.
- *Flight deck design issues*: In general the respondents were happy with the overall design in the automated cockpit. Issues were raised, however, regarding the interpretation of flight instrument displays and unnoticed events in map shift (loss of accuracy in navigational displays). Rudisill (1995) suggests that transition training for new ‘glass’ pilots should emphasise self-discipline and vigilance in monitoring raw data information.
- *Understanding how to use automation*: The general comments from participants regarding the integration and use of automation elements was positive. Some issues raised in this respect concerned pilots’ lack of knowledge about the intended behaviour of the aircraft in certain modes of flight. Pilots should have the ability to disconnect the automation and take manual control in the event of adverse aircraft behaviour in critical phases of flight to mitigate uncertainty (Rudisill, 1995).
- *Crew coordination and personal issues*: Respondents commented that ‘automation may reduce workload in low workload flight phases and may increase workload in high workload flight phases; also, workload may be increased dramatically during abnormal situations and failures’ (Rudisill, 1995, p. 290). It was also noted that crews were affected by boredom and complacency during periods of low workload. Again, crew discipline and improved systems knowledge helps to minimise this kind of problem.

Other studies conducted by Mosier, Skitka, Heers and Burdick (1998) have found extensive evidence that the advanced flight deck and the extensive use of automation have created an environment of automation bias and flawed heuristics (as a short-cut to decision-making, a symptom of complacency) which may threaten safety. As early as the 1990’s, research conducted by Parasuraman, Molloy and Singh (1993) and Parasuraman and Riley (1997), identified the need to optimise pilot workload in order to reduce boredom and mitigate the consequences of complacency. The identification of workload optimisation is important on an advanced flight deck. In an attempt to evaluate the differences in workload between pilots flying traditional analogue aircraft versus pilots operating modern flight deck aircraft, Laudeman and Palmer (1992), found a significant difference in aircrew performance. They indicated that those pilots that prefer to make use of the automated features in modern flight decks experienced higher workloads than those in traditional cockpits as well as those in automated flight decks that opt not to use the automation. Damos, John and Lyall (2005) also examined how the frequencies of 23 activities varied as a function of cockpit automation. The study examined general 'house-keeping' activities and communication, as well as flight path control, which may be regarded as one of the most fundamental factors in reducing aircraft accidents and incidents. That is, maintaining the correct flight profile is what keeps an aircraft in the correct (safe) three-dimensional space. Human factor errors emerge when the pilot has to cope with and integrate an excessive number of sources of information. Paradoxically, behavioural errors can also occur when the workload is too low. Workload conflict appears to be a problem that contributes to human error on the advanced flight deck (Kantowitz & Casper, 1988).

Because of the concerns of the effects of advanced automation on pilots behaviour, the United Kingdom (UK) Civil Aviation Authority (CAA) requested the RAF Institute of Aviation Medicine 'to assess the effects of advance automation on UK pilots in order to identify possible problems' and to research the 'opinions and attitudes of UK pilots to advance flight deck automation' (James et al., 1991a, p. 3.2). The researchers developed a questionnaire that included 78 items to assess pilots' opinions regarding aircraft automation. Ten of the items were related to general attitude towards aircraft automation and 68 items addresses several human factor concerns and automation issues on advanced flight decks. This included design, reliability, flight management system input, output and feedback, skills, training, crew interaction, monitoring and procedures, workload, and overall impressions. All the items of the questionnaire comprised of two statements defining opposite viewpoints with a Likert type scale, from 1 to 5 between them, indicating grades of opinion.

James et al. (1991a) used the questionnaire to survey the opinion of UK commercial pilots towards advanced flight deck automation. The attitude survey was distributed to all UK licensed pilots (approximately 11000). 1372 questionnaires were returned of which 572 were usable. Principle components analysis with varimax rotation of the responses on the 68 items identified four main factors. These were Understanding/Mastery, Workload, Design, and Skills. The four factors accounted for 31.48% of the explained variance (James et al., 1991a, pp. 3.2-3.5). Understanding/Mastery consisted of 'comprehension, expertise, knowledge and use of the system.' Workload entailed 'workload, demand, stress and task efficiency.' Design referred to 'ergonomic efficiency, design and displays' and Skills encompassed handling skills, crew interaction, and self-confidence' (p.3.5). Unfortunately the authors did not report on the reliability of the four factors, neither did they provide a list of all the significant items that signify the rotated factors (James et al., 1991a, b; McClumpha, James, Green, & Belyavin, 1991).

Singh, Deaton and Parasuraman (2001) 'developed (sic) a scale to assess pilot attitudes towards cockpit automation'. They used 30 items of the original questionnaire of the James et al. (1991a) survey. These items included the first 10 general attitude items and 20 items associated with the humanfactors and automated systems. Singh et al., (2001) used both positive and negative statements in the 30 item questionnaire. The favorable statements were scored on a scale ranging from strongly disagree (1) to strongly agree (5) and the unfavorable statements were scored on a scale ranging from strongly agree (1) to strongly disagree (5) (Singh et al. 2001, p. 208).

The questionnaire was administered to 170 pilots at Embry-Riddle Aeronautical University. 163 pilots with experience of advanced automated aircraft responded on the survey of which 111 completed the questionnaires satisfactory. Principal component analysis with varimax rotation of the responses on the 20 item section of the questionnaire revealed the presence of six factors with eigen values greater than one (Singh et al., 2001, p.208). These six factors, that accounted for 58, 3% of the explained variance, were named: workload (5 items with loadings of 0.58 to 0.77); design (5 items with loadings of 0.35 to 0.74); skills (4 items with loadings of 0.37 to 0.71); feedback (3 items with loadings of 0.63 to 0.70); reliability (4 items with loadings of 0.33 to 0.82); and self-confidence (3 items with loadings of 0.31 to 0.85). The reliability of the six factors was also computed using coefficient alpha which ranged from 0.75 to 0.98. Although the Singh et al., (2001) reported satisfactory reliability coefficients the inclusion of three items that cross-loaded on more than one factor are highly questionable. If these three items were omitted from the last two factors, both reliability and self-confidence would not have been included in the underlying factor structure of the questionnaire. According to Tabachnick and Fidell (2007, p. 646) the interpretation of factors defined by only one or two variables is not feasible.

Further research into human-automation interaction in today's advanced automated aircraft remains important to identify the core human factor concerns and automation issues related to current flight deck automation. The present study is an effort to aid in the identifying and describing specific areas of pilots concern regarding their performance in a highly advanced automated environment and their opinion about advanced automation systems. This article builds on the research of James et al. (1991a) to extend our knowledge and understanding of pilots' attitudes towards aircraft automation. The present study, however, differs from the preceding study in several ways. First, it examined the perception of airline pilots from a single South African airline. Second, all the pilots in the sample operate advanced third and fourth generation automated aircraft -- the state of the art technology in the industry today. Third, it focused on specific variables (individual and situational) that may account for variance in pilots' perception towards flight deck automation. New items were generated and a number of the original items from the survey of James et al. (1991a) were adapted to ensure relevance to the operational procedures and the types of aircraft the airline operates.

The primary objective of part 1 of this study was therefore to construct a valid and reliable instrument to measure the perceptions of airline pilots towards the core automation issues linked with operating advanced automated aircraft and to offer psychometric evidence for such a measure, termed the Automation Attitude Questionnaire (AAQ).

.

Research design

Research approach

In order to achieve the study objective a quantitative research approach based on the positivist paradigm was followed. A survey was conducted, using a structured questionnaire to collect the research data from a purposive sample of airline pilots. The data were analysed in accordance with the associational design as suggested by Field (2005, pp. 107, 619, 667). The associational design was used to establish the correlation between items scores on the questionnaire. The inter-correlation coefficients were employed to identify the underlying dimensions or factor structure of the questionnaire and to calculate the internal consistency and unidimensionality of the factors.

Research method

Participants

The research group represented a purposive sample of 262 current airline pilots at a major South African carrier operating both Airbus and Boeing type aircraft. Biographical information was elicited from all the participants in the first section of the questionnaire. The biographical characteristics of the participants are summarized in Table 1. Of this group, 245 were male pilots and 17 were female pilots. The small proportion (6.5%) of female participants was due to the fact that women have only recently started choosing professional flying as a career option. These numbers reflect the current low proportional representation of female pilots (6.1%) engaged in commercial aviation in South Africa (SACAA, 2007).

The sample ranged from lower entry pilots (in-flight relief crew) to high level pilots (senior training captains on long-range flights). It also represented diversity in terms of the type of aircraft flown, pilots' age and level of experience. 35.5% of the respondents had flown Boeing and 63.4% had flown Airbus type aircraft. The participants' ages ranged from 25 to 65 years (a spread of 40 years). There mean age was 44.14 years (SD = 9.556). The respondents' number of years of flying experience ranged from between 4 years and 46 years, with an mean of 23.73 years (SD = 10.373). The mean number of flying hours of the sample was 12 231 hours (SD = 5 636). The mean digital flight hours logged by the sample was 4 691.13 hours (SD = 2 530.004). The total digital flying time logged was expected to be significantly lower than the total flying time, as the carrier only began to operate modern automated aircraft in the last ten years or so. Only 24.9% of the respondents had any university level education.

Table 1.

Biographical Data of Respondents.

VARIABLE	FREQUENCY	PERCENTAGE
GENDER		
Male	245	93.5%
Female	17	6.5%
POSITION		
Dedicated in-flight relief pilot	16	6.1%
Co-pilot (Short Range)	60	22.9%
Co-pilot (Long Range)	49	18.7%
Captain (Short Range)	48	18.3%
Captain (Long Range)	53	20.2%
Training Captain (Short Range)	11	4.2%
Training Captain (Long Range)	18	6.9%
Other	5	1.9%
AGE		
25 – 35 years	59	22.5%
36 – 45 years	88	33.6%
46 – 55 years	67	25.6%
56 – 65 years	48	18.3%
LEVEL OF EDUCATION		
High school	163	62.5%
Diploma	33	12.6%
Bachelors degree	40	15.3%
Post Graduate	25	9.6%

INITIAL FLYING TRAINING		
Military	131	50.0%
Cadet	21	8.0%
Self (Part-Time)	72	27.5%
Self (Full Time)	37	14.1%

EXPERT YEARS		
4 to 15 years	63	24.0%
16 to 25 years	89	34.0%
26 to 35 years	65	24.8%
36 to 46 years	44	16.8%
Missing	1	0.4%

TOTAL DIGITAL FLYING TIME LOGGED		
0 to 2 000 hours	33	12.6%
2 001 to 3 000 hours	53	20.2%
3 001 to 4 000 hours	46	17.6%
4 001 to 5 000 hours	48	18.3%
5 001 to 6 000 hours	20	7.6%
>6 001 hours	60	22.9%
Missing	2	0.8%

TOTAL FLYING TIME LOGGED		
1 500 to 7 900 hours	65	24.8%
7 901 to 11 200 hours	69	26.3%
11 201 to 16 000 hours	56	21.4%
16 001 to 27 000 hours	69	26.3%
Missing	3	1.1%

Measuring Instrument

To identify the core human factor issues related to current flight deck automation and to assess airline pilots' perceptions of these issues, a measuring instrument called the Automation Attitude Questionnaire (AAQ) was constructed. Various research output in the field of flight deck automation were considered as points of departure in constructing the AAQ. The Items for the AAQ were generated by analysing the fundamental framework of research undertaken by Wiener (1989) and studies conducted by Funk and Lyall (2000), James, et al. (1991a, b) and Sherman et al., (1997).

The item pool of the initial AAQ included 85 items. Thirty three of these items were firstly selected and adopted from the 78 items of the attitude survey developed by James, et al., (1991a,b). Secondly, a further 35 items were extracted from the survey and adjusted to ensure clarity and relevance in the context of the South African airline that participated in the current study. Afterwards, these items were added to the AAQ. Thirdly, after discussions with experts and a further analysis of the literature, 17 new items were generated and were included in the questionnaire. Each of the 85 items of the initial AAQ presented one statement that covered various domains that encompass automation training, flying skills, workload, ergonomics, automation performance etc. All the statements (except for the biographic variables) were rated on a seven-point Likert-type scale to measure the perceptions of respondents at an approximate interval level. Unfavourable statements were scored on a scale ranging from strongly agree (1) to strongly disagree (7). The favourable statements were reverse coded to produce a measure where high scores indicated positive perceptions and low scores resulted in a more negative perception of automation.

In its final form, the preliminary AAQ consisted of three sections. Section 1 related to the pilots' biographical information. Section 2 consisted of the 85 items related to pilots' perceptions, opinions and behaviour regarding automation. This second section attempted to determine the core human factor issues and pilot concerns related to flight deck automation. An additional part, section 3 was added to gain qualitative input from respondents. Participants were given the opportunity to comment, either positively or negatively on operating highly advanced automated aircraft.

Research procedure

A list of all the airline pilots employed at a large South African carrier was obtained from the organisation's human resources department. Permission was granted from the executive and chief pilot of the particular company to distribute the questionnaires to the entire pilot population in their employment. A total of 800 questionnaires were distributed on an individual basis via a box-drop method.

In order to maximise the response rate a cover letter with the endorsement from management was attached to each questionnaire. The cover letter also stated the purpose of the research and further stressed voluntary participation and anonymity. Anonymity was ensured by eliminating the need to provide a name on the questionnaire. The completed questionnaires were collected manually from a dedicated collection box. A total of 262 (33%) usable questionnaires were received. According to Tabachnick and Fidell (2007), this number of responses was adequate for an exploratory factor analysis.

Statistical analysis

The main statistical analyses for the study were accomplished through the utilisation of the Windows Statistical Programme for the Social Sciences (SPSS) version 15. Exploratory factor analysis (EFA) was used to explore the internal structure and validity of the AAQ. EFA was carried out by means of principal axis factoring and rotated using the promax procedure with Kaiser's normalization to obtain an oblique generated factor solution for the AAQ. To assess compliance with the distribution requirements, Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy were used. In order to determine the number of significant item factors, Kaiser's criterion, Horn's parallel analysis and Catell's scree-plot were used (Tabachnick & Fidell, 2007). According to Hayton, Allen and Scarpello (2004), parallel analysis provides the most accurate estimate of the number of true factors in a complex dataset. The internal consistency of the AAQ was assessed by calculating the Cronbach Alpha coefficient for each factor. Item-reliability index of the individual items were calculated to establish whether the items contributed to the underlying construct of the factors (Gregory, 2004), and the average correlations between the items of each factor were calculated to examine the homogeneity and unidimensionality of the retained factors (Cortina, 1993; Clark & Watson, 1995). Frequencies and distributive statistics were used to describe the characteristics of the sample and to analyse the distribution (mean, standard deviations, skweness and kurtosis) of the responses.

Results

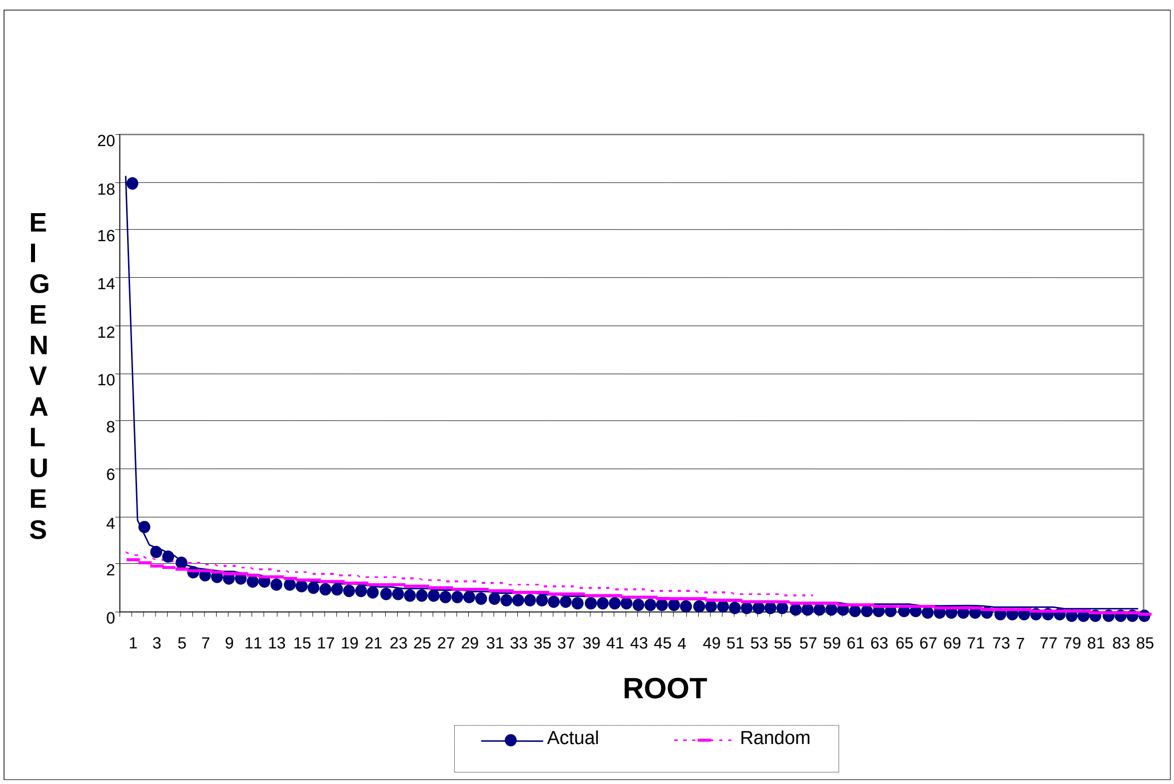
Exploratory Factor Analysis

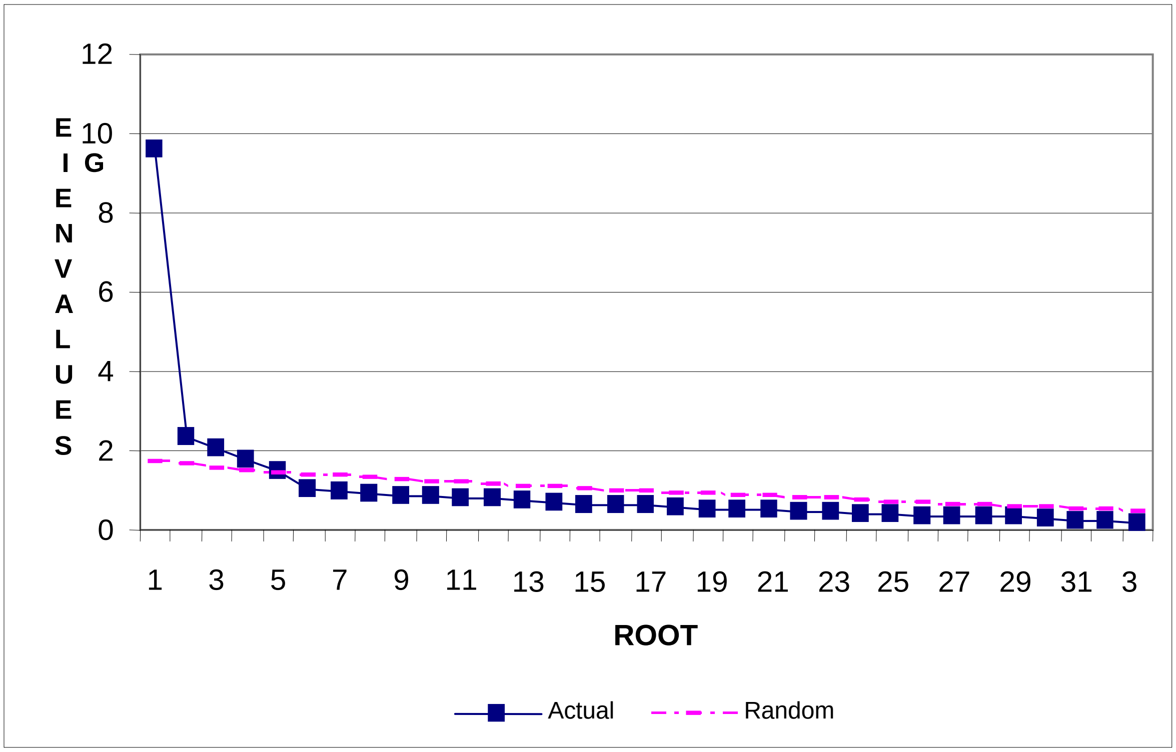
The Exploratory Factor Analysis (EFA) was carried out by means of principal axis factoring, and was rotated using the promax procedure ($\kappa = 4$) with Kaiser's normalisation to an oblique solution. This allowed the researchers to seek the lowest number of factors that account for the common variance in the set of 85 variables. In the first round of EFA, the 85 items of the AAQ were inter-correlated and rotated to form a simple structure by means of the promax rotation. Owing to the size (85 X 85), the inter-correlation matrix is not reported in the study. Based on Kaiser's (1961) criterion (eigenvalues larger than unity) 25 factors were postulated. The 25 factors explained 67.79% of the variance in the factor space of data. The factor analysis yielded more factors in the real test space than was expected. This is probably due to the presence of differentially skew items, as described by Schepers (1992). However, the results of Horn's parallel analysis and the scree-plot presented in Figure 1 confirmed that there were five significant constructs in the dataset. Parallel analysis indicated a break in the scree-plot between roots six and five. However, the curve of the eigenvalues of the random data set (the broken line) intersected the curve of the eigenvalues for the real data (the solid line) at root six. To avoid under-factoring, it was decided to include all the items of the six factors in the second round of EFA.

The items included in the six factors were first scrutinised; and the items which had factor loadings lower than 0.35 were omitted. A total of 33 items were retained and were subjected to the second round of EFA with promax rotation. The Kaiser-Meyer-Olkin (KMO) test for measuring sampling adequacy and Bartlett's test of sphericity displayed satisfactory results. Both diagnostic tests confirmed that the data was suitable for factor analysis. The calculated KMO value of 0.902 was greater than 0.7. Bartlett's test of sphericity [χ^2 (528) = 3470.758, p<0.01] confirmed that the properties of the correlation matrix of the item scores were suitable for factor analysis.

Six factors with eigenvalues higher than one were extracted in the second round of EFA. The six factors explained 55.297 percent of the total variance in the data. However, an inspection of the results of the parallel analysis presented in Figure 2, a five factors solution seemed more appropriate. Only one noteworthy item with a loading of 0.369 was associated with Factor Six. According to Tabachnick and Fidell (2007, p. 646), the interpretation of factors defined by only one or two variables is 'risky', under even the most exploratory of factor analyses. Consequently, Factor Six was disregarded. This resulted in a 32 item questionnaire that measure five factors related to flight deck automation. Of the 85 items included in the preliminary AAQ, 13 of the 33 original items and 13 of the 35 adjusted items from the James, et al. (1991a) survey, and five of the 17 new items were retained.

Figure 1.
Scree plot of the actual and the random data of 85 factors.





The factor loadings and corrected Item-total correlation of the items in each of the five factors of the AAQ are summarized in Table 2. The corrected item-total correlation of each item in the five factors was satisfactory and comply with the criteria suggested by DeVellis (2003) and Field (2005). DeVellis (2003) views an item with an item-total correlation of more than 0.20 as generally acceptable to be included. Field (2005) however, suggested that if an item-total correlation is less than 0.3, that a particular item should not be included as a variable in a scale. The values of the corrected item-total correlation in the five factors were all above 0.3. The percentage variance, sums of squared loadings, squared multiple correlations and factor correlations are reported in Table 3.

Table 2.

The factor loadings and corrected item-total correlation of the items that define the five factors of the AAQ.

Factor and Relevant Items	Factor Loading	Corrected Item-total Correlation
Factor 1		
Q38. I'm often confused about why the aircraft's automatics respond in the way it does.	0.831	0.724
Q36. I am often surprised by the aircraft's response to my FMS inputs.	0.816	0.722
Q40. I often tend to question the output from the automation system.	0.624	0.638
Q41. I find myself trying to guess what this aircraft is going to do next.	0.610	0.610
Q23. In the event of a partial system failure, it is never obvious which part of the automatic system failed.	0.567	0.475
Q37. I feel that the amount of feedback I get from the automatics is excessive.	0.557	0.557
Q42. The feedback I get in response to my inputs is usually too slow.	0.546	0.512
Q39. Even after receiving adequate feedback from the system, I still won't correct my fault.	0.433	0.519
Factor 2		
Q56. I think that there should be more simulator training for the conversion onto this aircraft.	0.831	0.698
Q55. The computer based-training was insufficient for me to fully understand this aircraft.	0.694	0.556

Q57. I feel that a lot more hours can be devoted to route training on this aircraft.	0.641	0.577
Q54. I think that there should have been a lot more classroom training for the conversion onto this aircraft.	0.631	0.644
Q58. There is insufficient recurrent training on this aircraft.	0.589	0.544
Q59. The training I received was inappropriate to line operations.	0.444	0.498
Q60. My transition onto this aircraft was extremely difficult.	0.367	0.443

Factor 3

Q78. I feel detached from the aircraft.	0.813	0.678
Q79. I feel exposed to risk by the automation.	0.745	0.683
Q77. The aircraft is always ahead of me.	0.671	0.642
Q80. Whenever I fly this aircraft, I feel a lot more stress then when I flew traditional aircraft.	0.605	0.594

Table 2 continued

Q75. The automation system greatly decreases my confidence as a pilot.	0.509	0.569
Q64. Automation impedes crew co-ordination.	0.495	0.651

Factor 4

Q73. The automation actually increases workload during critical phases of flight.	0.797	0.641
Q72. In the event of a flight plan change, the ‘heads-down’ time required is much more than in traditional flight decks.	0.733	0.591

Q69. I've noticed that there is much more ‘heads-down’ time in this cockpit.	0.575	0.465
Q71. It is very difficult for the crew to maintain a good look-out when flying this aircraft.	0.567	0.583
Q74. In general the overall workload on this flight deck has increased.	0.524	0.528
Q70. The procedures used to operate this aircraft don't suit it at all.	0.367	0.494
Factor 5		
Q16. I find that the aircraft automatics are extremely unreliable.	0.646	0.540
Q13. The displays in my aircraft make very poor use of colour.	0.590	0.438
Q17. The level of reliability and redundancy of the automatics is insufficient to conduct extended range operations.	0.522	0.445
Q14. I'm extremely unhappy with the set-up of the displays in my aircraft.	0.500	0.424
Q21. If the automatics fail, most of the time I don't try to restore the system.	0.421	0.404

Table 3.
Eigenvalues, Percentage Variance, Sums of Squared Loadings, Squared Multiple Correlations and Factor Correlations of the Five Factors of the AAQ.

Factors	**1**	**2**	**3**	**4**	**5**
Eigenvalues	9.577	2.315	2.049	1.780	1.473
Percentage Variance	29.022	7.014	6.210	5.393	4.463
Sums of Squared Loadings (SSL)	7.071	5.453	6.797	5.450	3.872
Squared Multiple Correlation (SMC)	0.991	0.974	0.977	0.980	0.930

Factor inter-correlation matrix

Factors	**1**	**2**	**3**	**4**	**5**
1. Comprehension	-	0.509	0.603	0.515	0.488
2. Training	0.509	-	0.501	0.419	0.250
3. Trust	0.603	0.501	-	0.569	0.412
4. Workload	0.515	0.419	0.569	-	0.351
5. Design	0.488	0.250	0.412	0.351	-

The five factor solution explains 51.102 percent of the total variance in the data. The five factors inter-correlated significantly with one another (r = 0.250 to 0.603). The strength of the correlations indicates that the five factors are closely related in measuring constructs related to flight deck automation. Although the relatively high inter-correlations suggest overlapping variability, the Squared Multiple Correlations (SMCs) of 0.930 to 0.991 between the item scores and the factor scores indicated that all the factors were sufficiently defined by the relevant items. The factor scores of the respondents were calculated by means of Bartlett's method as described in Tabachnick and Fidell (2007, p. 651). Scales were created for each factor and these were labeled according to the general content of their significant related items. The five factors or core automation issues linked with operating advanced automated aircraft were labeled Comprehension, Training, Trust, Workload and Design respectively.

The reliability of the five factors of the AAQ

The reliability of the factors of the Automation Attitude Questionnaire was determined by making use of Cronbach's coefficient alpha (Field, 2005). The average mean correlations between the items of each factor were also calculated to examine the internal homogeneity and unidimensionality of the five factors (Cortina, 1993; Clark & Watson, 1995). The means, standard deviation, skewness and kurtosis, average mean correlations and Cronbach's Alpha for the five factors are provided in Table 4.

Table 4.
Descriptive Statistics and Reliability of the Factors of the AAQ (n=262).

Descriptive Statistics	Compre hension	Training	Trust	Workload	Design
Mean	46.267	38.821	37.798	32.034	31.053
SD	7.079	7.401	4.606	6.224	3.992
Skewness	-1.265	-0.675	-1.584	-0.728	-2.085
Sk error	0.150	0.150	0.150	0.150	0.150
Kurtosis	2.124	0.011	3.323	0.302	7.276
Ku error	0.300	0.300	0.300	0.300	0.300
r(Mean)	0.59	0.56	0.63	0.55	0.45
Alpha	0.84	0.82	0.85	0.79	0.70

According to Table 4, the Cronbach alpha coefficients for the five factors of the AAQ were satisfactory. Compared to the guideline for alpha ≥0.70, recommended by Nunnally and Bernstein (1994), the alpha coefficient for the five factors yielded acceptable values (F1 =0.844; F2 =0.817; F3 =0.845; F4 =0.786; F5 =0.700). Furthermore, none of the items, if deleted, increases the internal consistency of a factor. All the mean inter-item correlations of the five factors were within the range of 0.15 to 0.50 as suggested by Clark and Watson (1995).The high mean inter-item correlations of 0.45 to 0.63 is probably the result of the specificity of the target constructs. According to Clark and Watson (1995), a much higher average inter-item correlation can be expected when one is measuring a narrow construct. The scores on the five factors of the AAQ appear to satisfy the requirements of homogeneity and unidimensionality and can be considered to be representative of the specific factor that they are assessing.

Discussion

The implementation of aircraft automation has for some time been the root of many debates within the aviation fraternity. The introduction of highly computerised flight deck technology has presented airline organisations with interesting human resource challenges. These challenges need to be met so as to maintain an efficient and optimal operational front. The perceptions of airline pilots with regard to flight deck automation issues has not yet been researched in South Africa and one of the challenges facing airlines is to determine what impact these perceptions have on successfully training and converting competent pilots to new advanced jet aircraft from traditional older generation aircraft (Barnett, 2005). The objective of part one of this study was therefore, to construct a valid and reliable instrument to measure the perceptions of airline pilots towards the core automation issues linked with operating advanced automated aircraft and to assess the psychometric properties of the measure and to refine the instrument.

A questionnaire named the Automation Attitude Questionnaire or AAQ was constructed to survey airline pilots' perceptions regarding automation issues linked with operating advanced automated aircraft. A total of 85 items were initially included in the AAQ. After two applications of exploratory factor analysis 32 of these items yielded a five-factor solution. The five factors showed adequate factorial validity, unidimensionality and reliability. The magnitudes of the factor scores off the items in each of the five factors were all larger then 0.35 with factor scores ranging from 0.36 to 0.83. The mean inter-item correlations ranged from 0.45 to 0.63 and the alpha coefficients from 0.73 to 0.85. These results provided sufficient support for the psychometric adequacy of the AAQ.
The five factors that associated with the core issues or demands of operating an automated flight deck or glass cockpit were labeled Comprehension, Training, Trust, Workload and Design.

Comprehension consisted of 8 items and includes issues such as how a pilot interprets and understands the capabilities, limitations, modes, and operating principles and functioning of the automated flight deck system. This factor includes pilot's competence to interpret the flight mode annunciator (FMA) and manage automation "surprises" (Parasuraman & Riley, 1997).

Training, the second dimension, was made up of seven items that refers to the training and learning required to get a pilot to an adequate standard or to the level needed to operate the automation system. The elements of this factor refer to quality time spent in classroom training, on simulator training, recurrent training, route training, line training and transition training on advanced aircraft.

The third factor was labeled Trust and includes six items that deal with the level of belief and assurance a pilot has in the performance of automated devices. It measured pilots' identification with the automation system; feelings of increased exposure to risk and stress due to automation, feelings that the aircraft is ahead of him or her; and being detached from the human-machine loop. A specific item of this factor also refers to the impedance in crew co-ordination due to system trust issues.

The fourth factor looked at perceptions of workload and includes six items. The primary issues covered in this factor are increases in workload during critical phases of flight. Elements of the workload factor consist of the amount of time spent instructing the automation computer via the flight management system (heads-down time) and thereafter having it accomplish a specific task correctly. Other elements also include the *procedures* required for safely operating the aircraft and the ability to maintain adequate situational awareness.

The fifth factor consists of five items related to the design characteristics and reliability of automation systems. This includes the ergonomic features and display design of the flight deck. Elements of the display design included the adequate presentation of accessible, useful, understandable and diagnostic visual and sound information, as well as the ease in utilising the information.

The five factors are closely linked to a number of human factor issues and concerns raised and cited by various authors in research publications and are also related to automation issues documented in international regulatory authority reports such as the FAA of the USA and the CAA of the UK. Elements of the present factors parallel those issues and demands associated with flight deck automation mentioned by Billings (1997), FAA (1996), Funk and Lyall (2000), James et al. (1991a), Mosier, et.al. (1998), Palmer (1995), Parasuraman and Riley (1997), Sarter and Woods (1994), Sarter, Woods and Billings (1997), Sherman (1997), Wiener (1989), Wood (2004). These human factor issues included the following variables: poor interface design; pilots' lack of understanding of the automated equipment; breakdown in attention and knowledge; demands in mode awareness and 'automation surprises'; uneven distribution of workload; over trust and decreased vigilance; pilot complacency and over-reliance on automation; loss of situational awareness; reduction of manual flying skills and proficiency; communication and coordination demands; and the need for new approaches to training.

Encouraging is the fact that the five factors identified in the present study also correspond with the ten prominent automation issues identified by Funk and Lyall, 1998 and Funk, et al. (1999). After intensive evidence based research, using various sources and criteria, Funk, *et al*., (1999, p. 120), listed the following five automation issues as the most important concerns that require solutions: ''understanding of automation may be inadequate; behavior of automation may not be apparent; pilots may be overconfident in automation; displays (visual and aural) may be poorly designed; and training may be inadequate".

A comparison of the results from this survey also indicates a strong commonality with the factors identified by James et al., (1991) and Singh et al., (2001). Workload, skills and design are common labels with understanding/mastery, self-confidence and comprehension sharing similar elements. Feedback, reliability and trust also appear to share common items. Overall the results indicate that common threads permeate pilot perceptions of automated flight decks and these are consistent over time. The results of this study resonates the capability of the AAQ to measure and assess airline pilots' perception of the most prominent issues and concerns in operating advance automated aircraft.

Practical application

The results of the statistical analysis of the responses on the AAQ suggest that the questionnaire is sufficiently reliable and valid to capture the present sample of airline pilots' perceptions of flight deck automation. Consequently aviation human factor specialists and aviation psychologists can use the instrument with confidence to gather valid and reliable data about automation perceptions held by airline pilots in South Africa. Understanding key concepts and fundamental issues associated with attitudes, perceptions and behaviour that exist within the sphere of advanced flight decks, has significant benefits for the aviation industry at large. A concise understanding of this topic will benefit airlines and other organisations to design and develop specifically targeted training material and to positively influence their pilots in accepting automation. However, elements that influence overall perceptions of automation may depend on the type of organisation, nature of flight training, flying experience, type of aircraft, computer literacy, operational position, etc. Further research should endeavour to identify those variables that may have an effect on the perceptions of airline pilots. In the second part of this research project the relationship between the biographical characteristics of the airline pilots' and their perceptions of automation were determined. The responses of the various pilot groups on the AAQ were compared to one another and where applicable correlated with their scores on the different factors. These results are offered in the part two of the article series.

Acknowledgements

This article is part of a research project on flight deck automation co-ordinated by Professor Dr. Leopold P. Vermeulen (The University of Pretoria).

References

Airbus A380 Cockpit. Retrieved March 29, 2007, from the World Wide Web: http://www. gillesvidal.com/blogpano/cockpit1.htm

Barnett, J. S. (2005).Training people to use automation: Strategies and methods. *Journal of Systemics, Cybernetics and Informatics*, 3(5), 73-76.

Billings, C.E. (1997). *Aviation Automation*. Mahwah, NJ: Lawrence Erlbaum.

Clark, L. A. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309-319.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98-104.

Crouch, T. D. (2003). *The Bishop's Boys: A Life of Wilbur and Orville Wright*. New York: W. W. Norton & Company.

Damos, D.L., John, R.S. & Lyall, A.E. (2005). The Effect of Level of Automation on Time Spent Looking Out of the Cockpit. *The International Journal of Aviation Psychology,* 9(3):303-314.

DeVellis, R. F. (2003). *Scale Development: Theory and Applications.* Thousand Oaks, CA: Sage.

Edwards, E. (1988). Introductory Overview. In E.L. Wiener & D.C. Nagel (Eds.), *Human Factors in Aviation,* (pp. 3-25). San Diego, CA: Academic Press.

Field, A. (2005). *Discovering Statistics using SPSS*. (2nd ed.). London: Sage.

FAA. (1996). *The interfaces between flightcrews and modern flight deck systems*. Human Factors Team, Washington, DC: Federal Aviation Administration.

Funk, K., Lyall, B. & Riley, V. (1995). Flightdeck Automation Problems. *Proceedings of the Eighth International Symposium on Aviation Psychology, (pp.*265-269). Columbus, OH: The Ohio State University.

Funk, K., Lyall, B., Wilson, J., Vint, R., Niemczyk, M., Suroteguh, C. & Owen, G. (1999). Flight Deck Automation Issues. *The International Journal of Aviation Psychology*, 9(2), 109-123.

Funk, K., Lyall, B. (1999). The evidence for flight deck automation issues. *Proceedings of the Tenth International Aviation Psychology Symposium Conference(CD-R)*. Columbus, OH: The Ohio State University.

Funk, K. & Lyall, B. (2000). A Comparative Analysis of Flight Decks with Varying Levels of Automation. *Final Report prepared for the FAA Chief Scientific and Technical Advisor for Human Factors,* (pp. 1-17). Washington DC: Federal Aviation Administration.

Gregory, R. J. (2004). *Psychological testing: history, principles, and applications,* (4th ed.). Boston. Pearson Education Group.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191-205.

Ishibashi, A., Kanda, N. & Ishida, T. (1999). Analysis of Aircraft Accidents by Means of Variation Tree. *Proceedings of the Tenth International Aviation Psychology Symposium Conference (CD-R)*. Columbus, OH: The Ohio State University.

James, M., McClumpha, A., Green, R., Wilson, P. & Belyavin, A. (1991a). Pilot Attitudes to Flight Deck Automation. *Proceedings of the Royal Aeronautical Society Conference on Human Factors on Advanced Flight Decks,* (pp. 130 – 158). London, UK: Human Factors Society.

James, M., McClumpha, A., Green, R., Wilson, P. & Belyavin, A. (1991b). Pilot attitudes to cockpit automation. *Proceedings of the Sixth International Symposium of Aviation Psychology,* (pp.192-198). Columbus, OH: The Ohio State University.

Kabbani, M.A. (1995). The Glass in the Cockpit – Cloudy or Clear? *Proceedings of the Eighth International Symposium on Aviation Psychology,* (pp. 64-67). Columbus, OH: The Ohio State University.

Kabay, M. E. (1996). Advanced Automated Flight Deck Issues. Retrieved April 01, 2007, from the World Wide Web: http://www.ncsa.com/articles/incidents

Kantowitz, B. & Casper, P. (1988). Human Workload in Aviation. In E. L. Wiener and D.C. Nagel (Eds.), *Human Factors in Aviation* (pp. 157-187). San Diego, CA: Academic Press.

Laudeman, I. V. & Palmer, E. A. (1992). Measurement of automation effects on aircrew workload. *In the Third Annual ASIA Program Investigator's Meeting*. Moffett Field, CA: NASA-Ames Research Center.

Lyall, B. & Funk, K. (1998). Flight Deck Automation Issues. In M.W. Scerbo & M. Mouloua (Eds.), *Proceedings of the Third Conference on Automation Technology and Human Performance*, (pp. 288-292). Norfolk, VA, March 25-28, Mahwah, NJ: Lawrence Erlbaum Associates.

McClumpha, A. J., James, M., Green, R. G., & Belyavin, A. J. (1991). Pilots'Attitudes to Cockpit Automation. *In Proceedings of the Human Factors Society 35th Annual Meeting,* (pp. 107–111). Santa Monica, CA: Human Factors and Ergonomics Society.

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.

NTSB, National Transportation Safety Board. 2009. *Accident and Incident Report for Part 121 Operators.* [Online] Available from: http://www.ntsb.gov/ntsb/AVIATION/ [Downloaded: 2009-03-17]. Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric Theory. (3rd* ed.). New York: McGraw-Hill.

Palmer, E. (1995). Oops, "It Didn't Arm." A Case Study of Two Automation Surprises. *Proceedings of the Eighth International Symposium on Aviation Psychology*, (pp. 227-232). Columbus, OH: The Ohio State University.

Paolo, R. (2001). *The Birth of Modern Science*. London: Blackwell.

Parasuraman, R., Molloy, R. & Singh, I. (1993). Performance Consequences of Automation Induced Complacency. *The International Journal of Aviation Psychology,* 3(1), 1-23.

Parasuraman, R. & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *The International Journal of Aviation Psychology,* 39(2):230-253.

Risukhin, V. (2001). *Controlling Pilot Error. Automation.* New York: McGraw-Hill.

Rudisill, M. (1995). Line pilots' attitudes about and experience with flight deck automation: Results of an international survey and proposed guidelines. *Proceedings of the Eighth International Symposium on Aviation Psychology,* (pp. 288-293), Columbus, OH: The Ohio State University.

SACAA. 2007. [Statistics per licence on gender. XLS] According to C. Lakay, (e-mail communication, July 2007; LakayC@caa.co.za)

Sarter, N.B. (1995). Knowing when to look where: Attention allocation on advanced automated flight decks. *Proceedings of the Eighth International Symposium on Aviation Psychology,* (pp.239-241). Columbus, OH: The Ohio State University.

Sarter, N.B. (1996). Human errors are symptoms of a mismatch between pilots, machines, and the operating environment. *ICAO Journal*, 51(8), 9-10.

Sarter, N.B., & Woods, D.D. (1992). Pilot interaction with cockpit automation: operational experiences with the flight management system. *The International Journal of Aviation Psychology*, 2 (4):303-321.

Sarter, N. B. & Woods, D. D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' mental model and awareness of the Flight Management System (FMS). *The International Journal of Aviation Psychology*, 4(1), 1-28.

Sarter, N.B., Woods, D. D. & Billings, C.E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of Human Factors & Ergonomics*, (2nd ed.), (pp 1-25), New York: John Wiley.

Schepers, J.M. (1992). *Toetskonstruksie: Teorie en Praktyk.* Johannesburg: RAU

Singh, I., Deaton, J. & Parasuraman, R. (2001). Development of a Scale to Assess Pilot Attitudes Towards Cockpit Automation. *Journal of the Indian Academy of Applied Psychology*, 27(1-2), 205-211.

Sherman, P.J. (1997). *Aircrews' Evaluations of Flight Deck Automation Training and Use: Measuring and Ameliorating Threats to Safety1*. Technical Report 97-22.
FAA Grant 92-G-017.

Sherman. P.J., Helmreich, R. L. & Merritt, A.C. (1997). National Culture and Flightdeck Automation: Results of a Multi-nation Survey. *The International Journal of Aviation Psychology,* 7(4), 311-329.

Skitka, L.J., Mosier, K.L., Burdick, M. & Rosenblatt, B. (2000). Automation Bias and Errors: Are Crews Better than Individuals? *The International Journal of Aviation Psychology,* 10(1):85-97.
Tabachnick, B.G. & Fidell, L.S. (2007). *Using Multivariate Statistics, (*5th ed.). Boston, MA: Allyn & Bacon.
The Boeing Company. (2009). Message Number: MOM-MOM-09-0063-01B. 04-March-2009.
Wiener, E. L. (1988). Cockpit Automation. In E.L. Wiener & D.C. Nagel (Eds.), *Human Factors in Aviation,* (pp. 433-461). San Diego, CA: Academic Press.
Wiener, E. L. (1989). *Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft*. NASA Contractor Report 177528, Moffett Field, CA, USA.
Wiener, E. L. (1993). Crew Coordination and Training in the Advanced Cockpit. *Cockpit Resource Management.* San Diego, CA: Academic Press.
Wood ,S. (2004). *Flight Crew Reliance on Automation*. CAA Paper 2004/10.
Research Management Department, Safety Regulation Group, Civil Aviation Authority, UK, Gatwick Airport South, West Sussex.

IDENTIFYING THE IMPACT OF NEXTGEN ON THE JOB OF AIR TRAFFIC CONTROL SPECIALISTS IN THE MID-TERM

Emily Baumann, MS
Kelley J. Krokos, PhD
Cheryl Hendrickson, PhD
American Institutes for Research
Washington, DC 20007

By leveraging existing and new technology, the Next Generation Air Transportation System (NextGen) is proposed to support significant increases in both capacity and efficiency of the National Airspace System (NAS). In the decade that has passed since NextGen was mandated by Congress, the Federal Aviation Administration (FAA) has made great progress in identifying the specific technologies, automation, and procedures (i.e., NextGen Drivers) necessary to support the desired increases. American Institutes for Research (AIR) describes its strategic job analysis, which was designed to identify the NextGen Drivers that are proposed to be implemented by the NextGen Mid-Term (2018), and to evaluate how the job of Air Traffic Control Specialists (ATCSs) would change as a result of their implementation. Implications for human resource systems such as pre-employment selection and training are discussed.

The Next Generation Air Transportation System (NextGen) is part of the VISION 100 – Century of Aviation Reauthorization Act (P.L. 108-176) mandated by Congress nearly a decade ago. The impetus of NextGen was the need to increase the capacity of the National Airspace System (NAS) to meet the predicted demand for US air travel in 2025 while simultaneously maintaining or improving reliability and safety. In order to accomplish significant increases in both capacity and efficiency, the implementation of NextGen will leverage existing and new technology, including satellite-based surveillance and navigation. Collectively, these new technologies and the automation, and procedures they support (labeled “NextGen Drivers”) are proposed to be used – either directly or indirectly – by Air Traffic Control Specialists (ATCSs). For example, Data Communications is a digital communication system that will provide ATCSs with the capability of directly communicating with pilots using a computer-based data entry system—assuming that the aircraft are properly equipped and that the message meets certain operational guidelines.

It is necessary to understand the nature of the impact of the NextGen on the job of ATCSs, so as to inform important processes including whether pre-employment selection or training for the job will need to be modified. With this purpose in mind, the Federal Aviation Administration (FAA) funded the American Institutes for Research (AIR) to develop a vision of the job of ATCSs as it is proposed to exist in the Mid-Term 2018 stage of the NextGen evolution. As described below, AIR researchers conducted a Strategic Job Analysis (SJA) to identify and define the job of ATCSs working in all three job options – Airport Traffic Control Towers (ATCTs), Terminal Radar Approach Control Facilities (TRACONs), and Air Route Traffic Control Centers (ARTCCs).

Methodology

An SJA is the process of defining a job that is being redesigned or created in terms of the work that employees will perform and the characteristics of workers that will be required to perform those activities (Schippmann, 1999). In the case of job redesign, it is especially important to take note of the impact that any system changes will have on the job and redefine the job in its anticipated form. As a result, the first step of the SJA was to update the job analysis data for the ATCS job as it is currently performed. AIR conducted a series of focus groups with subject matter experts (SMEs) including ATCSs, managers and supervisors, and trainers and instructors from both the FAA Academy as well as FAA facilities. The goal of the focus groups was to update the lists of ATCS Tasks required to perform the job and the Knowledge, Skills, Abilities, and Other Personal Characteristics (KSAOs) required to perform the job successfully. In addition, AIR developed a comprehensive and current list of the Tools and Equipment used by ATCSs.

Next, researchers identified the NextGen Drivers that would likely impact the ATCS job by 2018. The Drivers were identified through a number of research activities. First, the NextGen research literature was reviewed, including NextGen Concept of Operations documents (FAA, April 2010), implementation plans (FAA, March 2010), human factors roadmaps, and functional requirements documents for specific technologies and automation. Next, focus groups and interviews with NextGen SMEs were conducted to get detailed information about NextGen plans, specific NextGen Drivers, and the likely impact of the Drivers on the ATCS job. These experts included systems engineers, human factors engineers, and aerospace engineers, and other individuals actively involved in conceptualizing NextGen or designing specific Drivers.

After identifying the NextGen Drivers that were proposed to be implemented by 2018, researchers examined and compared the Drivers to the FAA's list of NextGen Operational Improvements (OIs). The FAA's OIs capture specific enhancements to processes associated with the progression of aircraft through the NAS. The goal was to assure that the FAA's OIs were captured in one or more of the identified Drivers. First, AIR determined which OIs would affect which job option(s). OIs that were not expected to affect the ATCS job in at least one job option were eliminated from further consideration. Next, AIR conducted research to determine which Driver(s) would support or operationalize each OI in each facility type to illustrate the relationship between OIs, the three ATCS job options, and the specific mechanism (i.e., NextGen Driver) that would enable the OI to be realized. In both steps, four researchers made these determinations independently and then discussed their individual assignments as a group until they reached consensus.

The final step of the SJA was to determine the impact of the NextGen Drivers on the ATCS job as it is performed in the three job options (i.e., ATCT, TRACON, and ARTCC). The goal was to identify the changes to the ATCS job and to draw conclusions regarding the impact of these changes on the pre-employment selection and training systems. To begin, researchers worked independently to identify if the implementation of a Driver would require new Tasks to be added to the existing lists; would require existing Tasks to modified or deleted from the list; and whether it would change how a Task would be conducted in 2018 (e.g., if a new Tool or Equipment would be added or if information sources would change). After completing the research independently, researchers met and discussed the results until they reached consensus. The process was the same for the KSAOs with the exception that researchers also determined whether a Driver would change a characteristic of KSAOs required in 2018 (e.g., if a KSAO would become more or less important or the required level of proficiency would change).

Results – Identification of NextGen Drivers

A complete list of the 27 NextGen Mid-Term Drivers identified in 2010 by AIR for each facility type is shown in Table 1 below. Drivers range from aircraft equipage supported by satellite technology (e.g., Automatic Dependent Surveillance-Broadcast) to new ATCS automation (e.g., Tower Flight Data Manager) with embedded decision support tools (DSTs) that will be added to the ATCSs' work environment to new policy and procedures (e.g., Integrated Arrival/Departure Air Traffic Control Service) consisting of new routes and new separation standards that ATCSs will be expected to follow. As seen in the table, AIR defines Drivers by facility type. For example, 4-D Wx Data Cube is included in the table three times, once for each facility type in which it is expected to be deployed. This method, while it allows AIR to describe the sometimes considerable differences in the impact of the Driver by facility type, results in 27 Drivers when there will be fewer actual core programs.

Table 1.
Crosswalk of NextGen Drivers by ATCS Job Option.

ATCT	TRACON	ARTCC
4-Dimensional Weather Data Cube	4-Dimensional Weather Data Cube	4-Dimensional Weather Data Cube
Airport Surface Detection Equipment-Model X		
Automatic Dependent Surveillance-Broadcast	Automatic Dependent Surveillance-Broadcast	Automatic Dependent Surveillance-Broadcast
		Conflict Resolution Advisories
Data Communications		Data Communications
	Flexible Airspace Management	Flexible Airspace Management
		High Altitude Airspace
	Integrated Arrival/Departure Air Traffic Control Service	Integrated Arrival/Departure Air Traffic Control Service
Integrated Arrival, Departure, and Surface	Integrated Arrival, Departure, and Surface	
	Optimized Profile Descent	Initial Tailored Arrivals
	Performance-Based Navigation	Performance-Based Navigation
Terminal Automation Modernization and Replacement, Phase 3	Terminal Automation Modernization and Replacement, Phase 3	
		Time-Based Flow Management Program
Tower Flight Data Manager		
Wake Turbulence Mitigation for Departures	Wake Turbulence Mitigation for Arrivals	

Results – Impact on Tasks, KSAOs, and Tools and Equipment

Results of the SJA revealed several effects of the NextGen Drivers on the ATCS job. In only a few cases, the implementation of the NextGen Drivers will require additional (new) job Tasks. For example, in ATCT and ARTCC, new tasks will be required to establish and terminate Data Communications. However, across job options, the results of the research showed that the job Tasks performed by ATCSs will change very little by 2018. That is, few additions, deletions, or modifications will need to be made to the existing Task lists. This is due in part to the nature of the NextGen Drivers; they are designed to help controllers do their job more efficiently rather than change *what* they do. ATCSs will have access to more information and also more accurate information, which will improve their situation awareness and decision-making. Although relatively few changes are called for in terms of changes to the existing Task lists, there will be a number of changes regarding *how* the job Tasks are performed. For example, new workstation automation will change how ATCSs perform their current job responsibilities by 2018. Finally, the implementation of the Drivers adds more decision branches that controllers must navigate to conduct various Tasks, which could lead to an increase in their mental burden. However, this increase could be offset by the introduction of DSTs designed to help controllers work more quickly and to allow them to focus on other more challenging aspects of their job. However, note that the net effect on workload is currently unknown.

In addition to having an impact on job Tasks, the Drivers will also have a minor impact on the requirements of individuals who perform the ATCS job. With regard to KSAOs, across job options and Drivers, the changes include the addition of two new Knowledges (i.e., *ATC Automation* and *Interoperability*), one new Skill (i.e., *Service Orientation*), and one new Other Personal Characteristics (i.e., *Technology Acceptance*). Training content for *ATC Automation* would include the evolution of ATC automation; risks associated with automation (e.g., over or under reliance on automation); benefits of automation (e.g., freeing of cognitive resources for use on other Tasks); automation design considerations including appropriate Task allocation to man and machine; and concepts associated with DSTs including the decision support tool – decision making tool continuum, evaluation strategies, and the concept of automation-based algorithms and the importance of understanding them. Training content for *Interoperability* would include how a specific Tool and Equipment interacts with other Tools and Equipment. This could include information regarding how data will be depicted on a display, how the system as a whole interacts with other existing systems, and if information will be displayed on more than one system.

One new skill was added: *Service Orientation* requires ATCSs to be skilled in providing service to properly equipped air carriers. For example, ATCSs will be able to provide certain options and routes to properly equipped aircraft that will not be available to others. Again, because actual Tasks performed by ATCSs will change very little, no new Abilities will be required. Finally, with regard to Other Personal Characteristics, *Technology Acceptance,* defined as the need for controllers to have positive attitudes towards, perceived usefulness of, and perceived ease of use of technology, will be added. While the Drivers will affect the relative importance of KSAOs, making some more important and some less important in terms of a specific Driver, the overall net effect is not known.

Finally, the Drivers will increase the number of Tools and Equipment in the ATCS work environment. More often than not, a new Tool or Equipment will be added to the ATCS workstation in the Mid-Term rather than replacing an existing one. This is especially significant in the case of the ATCT environment, which has traditionally relied on an out-the-window view to gather information rather than looking down at a radar display as is done in the other job options. The significant increase of Tools and Equipment in ATCTs will increase the heads down time of controllers. This increase in heads down time has grown over time and marks another step along the automation continuum toward full automation. While the net effect of the increased heads down time is unknown, it should be investigated further.

While determining the impact of NextGen Drivers on the ATCS job, AIR identified 19 potential risks associated with the implementation of the Drivers. The risks were grouped into four categories: risks associated with implementation of technology (e.g., Improper Allocation of Tasks to Automation; Mixed Aircraft Equipage); risks associated with the implementation of new NextGen policies and procedures (e.g., Best Equipped, Best Served (BEBS); risks associated with the new NextGen work environment (e.g., Change in Culture; More Dynamic Work Environment); and risks associated with individual controller job performance (e.g., Improper Reliance on Automation or Procedures; Lack of/Inadequate Training). The impact of the risks varies, with some risks being associated with only one or two Drivers (e.g. Loss of Party Line Information), and others being associated with all Drivers (e.g. Technology Development and Maturation). Note that these are potential risks only that may or may not materialize depending on a number of currently unknown factors. Finally, they are not meant to constitute an exhaustive list of all NextGen risks but rather are a summary of the issues based on this research.

Results – Impact of NextGen on Human Resource Systems

As no new Abilities and only one new Other Personal Characteristic will be required as a result of NextGen by 2018, the impact on pre-employment selection is limited. No major revision of the AT-SAT pre-employment selection test battery was recommended. One area for future consideration is whether the current test battery captures the required Abilities and Other Personal Characteristics in the proper proportions. That is, pre-employment selection tests typically sample from the domain of all required Abilities and Other Personal Characteristics in proportions that mirror the importance of those characteristics to performing the job. The impact of NextGen on the required proportions is unknown and should be investigated.

Although the implementation of NextGen Drivers has limited impact on the pre-employment selection system, the impact on training is substantive. This is because of the large number of Knowledges and Skills that are proposed to be affected by the Drivers. Significant investment in developing new training content and training developmental and currently certified ATCSs to prepare them to successfully perform their job in the year 2018 will be required. The Strategic Training Needs Analysis (STNA) conducted by AIR as a follow on task to the SJA estimates in much more detail the impact on training, including the resources required to development, implement, evaluate, and maintain NextGen ATCS training.

Summary and Next Steps

AIR's research suggests that although changes lie ahead for the ATCS job, the job responsibilities for ATCSs will remain largely the same in 2018. ATCSs are provided with additional job aids and improved information and tools, but they remain responsible for separation of aircraft in the Mid-Term. The FAA's current pre-employment selection test battery will be suitable for selecting ATCSs to perform in 2018. However, significant additional training will be required to ensure that ATCSs are fully prepared to perform their jobs in 2018.

With regard to next steps, it was recommended that the FAA begin now to identify the processes that need to be put into place to support these changes including determining how to develop high quality standardized training, and identifying and procuring the resources required to make these modifications. Note that AIR's subsequent STNA completed in late 2012 provided critical information that the FAA needs to begin addressing this issue. In addition to addressing the training requirements for ATCSs, the impact of NextGen on other jobs should be considered. For example, additional work should endeavor to develop NextGen Job Descriptions for other air traffic control positions that interact directly with the line controller such as the traffic management unit (TMU) coordinator, with whom the line controller interacts. Another job category that will be affected is the technical operations (TechOps) job as TechOps employees are responsible for installing and maintaining the Drivers and/or their enabling technology. Although the potential risks identified by AIR are not a comprehensive list of all potential risks, research should be conducted to determine their likelihood and to develop potential mitigations. Finally, it should be noted that NextGen is a fast-moving and evolving initiative; these results are now somewhat outdated. It is recommended that the impact of NextGen on the ATCS job be revisited. Although the results suggest that few if any changes in responsibility will occur by 2018, this may not be the desired result. That is, if it is the desire of the FAA that the job will change by 2018 or beyond (i.e., the Far Term), then a different type of analysis may be required.

Acknowledgements

The American Institutes for Research would like to thank Mr. Dino Piccione, Technical Lead for Human Factors Air Traffic/Technical Operations Research and Ms. Barbara L. Wilper, Scientific and Technical Advisor for Human Factors, both of the Federal Aviation Administration's (FAA's) Human Factors Research and Engineering Group. In addition to sponsoring this research, they also provided significant technical guidance and support.

Managers from the FAA's Air Traffic Technical Training and Oversight group also assisted. Ms. Greta Ballard, Mr. Daniel Lacroix, Mr. Gregory Sanders, and Mr. Mark Marchese procured both information and access to subject matter experts, and participated in focus groups and other reviews.

Numerous other individuals made significant contributions. In particular, NextGen and air traffic control subject matter experts from the FAA, the National Air Traffic Controllers Association (NATCA), and numerous contracting organizations participated in interviews that informed the ideas presented.

The views expressed in this report are those of the authors. They do not necessarily reflect the views of the Department of Transportation, the FAA, NATCA, or any other organization.

References

Schippmann, J. S. (1999). *Strategic job modeling: Working at the core of integrated human resources.* Mahwah, NJ: Erlbaum.

U.S. Department of Transportation, Federal Aviation Administration (2010, March). *NextGen implementation plan*. Washington, DC: NextGen Integration and Implementation Office.

U.S. Department of Transportation, Federal Aviation Administration (2010, April). *NextGen mid-term concept of operations for the National Airspace System [Initial Coordination Draft]*. Washington, DC: Federal Aviation Administration.

NEXTGEN OPERATIONAL IMPROVEMENTS: WILL THEY IMPROVE HUMAN PERFORMANCE

Bettina L. Beard and James C. Johnston
NASA Ames Research Center
Moffett Field, CA, USA
Jon Holbrook
San Jose State University
Moffett Field, CA, USA

Modernization of the National Airspace System depends critically on the development of advanced technology, including cutting-edge automation, controller decision-support tools and integrated on-demand information. The Next Generation Air Transportation System national plan envisions air traffic control tower automation that proposes solutions for seven problems: 1) departure metering, 2) taxi routing, 3) taxi and runway scheduling, 4) departure runway assignments, 5) departure flow management, 6) integrated arrival and departure scheduling and 7) runway configuration management. Government, academia and industry are simultaneously pursuing the development of these capabilities. For each capability, the development process typically begins by assessing its potential benefits, and then progresses to designing preliminary versions of the tool, followed by testing the tool's strengths and weaknesses using computational modeling, human-in-the-loop simulation and/or field tests. We compiled research studies of the tools, assessed the methodological rigor of the studies and served as referee for partisan conclusions that were sometimes overly optimistic. Here we provide the results of this review.

The FAA's Next Generation Air Transportation System (NextGen) proposes to modernize the U.S air traffic system by deploying advanced technology with the aims of streamlining equipment, consolidating common operational tasks and facilitating human management of traffic operations. The FAA has identified specific capabilities, referred to as Operational Improvements (OIs) that will be gradually phased into operations. The OIs of interest here are those classified as Decision-Support Tools (DSTs) and procedures for airport tower personnel. They are:

- Departure Metering
- Taxi Routing & Scheduling
- Departure Runway Assignment
- Runway Scheduling
- Departure Flow Management
- Integrated Arrival/Departure, and
- Runway Configuration Management.

In 2010, the Government Accountability Office (GAO) recommended that the FAA "identify clear goals for the performance of these capabilities or … settle on a set of metrics for measuring their performance relative to any goals." In response, the FAA identified 5 metrics: capacity, efficiency, predictability, safety and environment (NextGen Implementation Plan, 2012). The FAA's Human Factors Research and Engineering Group would like for this set of metrics to also include measurements of human performance. NASA's Human Factors Research & Technology Division at Ames Research Center has entered into an Interagency Agreement with the FAA group to help define a human performance metric(s). We refer to the project as the Human Performance Budget.

NASA has approached the development of a Human Performance Budget in several ways. First, we asked a panel of human factors experts to rate whether the proposed OIs could have a positive or negative impact on 23 human performance metrics (Beard, Parke, Holbrook & Oyung, in preparation). The experts' ratings indicated that overall the OIs could, if implemented appropriately, positively influence team situation awareness and coordination. One of the most useful capabilities to offer the controllers would be automated support for conformance monitoring. According to the expert ratings the greatests risks were the potential for decay of controller skills and knowledge, potential difficulties with tools requring controllers to continuously monitor for the occurrence of a specified target or event and potential problems with meeting training standards using the new tools.

NASA then asked how thoroughly the OIs address current tower safety problems. We extracted and analyzed over 200 Aviation Safety Reporting System (ASRS) reports submitted over a 5-year period (Holbrook, Puentes, Stasio, Jobe, McDonnell & Beard, 2011). We found that the majority of the reports dealt with potential safety threats to runway operations, a concern that could potentially be addressed by OIs aimed at improving controller situation awareness during runway operations. Two issues extracted from the ASRS, oganizational climate and inadequate supervision, are not addressed by any of the current OIs.

NASA also asked the user community what they need. In a survey involving over 125 tower controllers, we asked to what extent the NextGen capabilities could help them in their job or to improve capacity, efficiency, flexibility, predictability and safety at their airport. Controllers indicated that departure metering from the ramp could be the most helpful tool but that enhanced information would be the most helpful enabler (Holbrook, Parke, Oyung, Collins, Gonter & Beard, 2013).

All three approaches to the Human Performance Budget point to the importance of keeping the controller in the loop and providing them with information to help them do their job. But it remains to be seen whether the advanced automation tools being developed will (1) improve controller and team situation awareness and coordination to aid in decision-making and (2) be mature enough for implementation by 2018.

Methods

Research articles that evaluated tower controllers' performance using NextGen systems were gathered and those articles judged to be of the greatest value were reviewed in considerable depth based on experimental design and relevance to NextGen. The goal was to evaluate whether the research and data analyses were experimentally rigorous and the conclusions valid. In a report to the FAA, Beard (2012) provides a more in-depth discussion of the literature review.

Research studies were classified based on their relevance to the seven capabilities listed in the Introduction (identified from NextGen OIs) and key enabling technologies required to realize the full potential of those capabilities. This report provides a brief description of the meta-analysis of Human-in-the-Loop (HITL) publications classified within the seven capabilities of NextGen. Research related to the key enabling technologies is evaluated elsewhere (Beard, Galeon & Parke, in preparation).

Results

Capability 1: Departure Metering at the Ramp

The need to reduce airport surface delays has been approached from different angles. Departure metering, prior to release from the gate or from the spot, is one approach that has considerable promise and has been implemented at several airports. The basic concept of departure metering is very simple: aircraft destined for the same runway are provided either gate release times or taxi times from the spot. This provides a way to sequence departing aircraft without the many drawbacks of a long physical queues (which impose costly time delay and fuel consumption penalties). This idea has been computationally shown to be advantageous both operationally and environmentally (e.g., Brinton & Lent, 2012; Simaiakis, Sandberg, Balakrishnan and Hansman, 2012).

The Surface Management System (SMS) is a system developed jointly by the FAA and NASA. An early field test focused on controller judgement (Atkins, Brinton, Walton, Arkind, Moertl & Carniol, 2003). Six years later, the tool was still found to provide unreliable pushback predictions (Monroe, 2009) because it did not, at the time, incorporate airport surveillance data.

At NASA Ames, an airport surface testbed based on SMS is being used to test the efficacy of tools for taxi and runway scheduling (Jung et al., 2010). With a predetermined sequence and reliable estimate of taxi times, each flight can be assigned pushback times calculated to allow unimpeded taxiing to the runway, followed by an immediate clearance for takeoff. Even though this concept has been tested in high fidelity simulation twice, it is still an immature concept. Results indicate that the concept does not appropriately take into account the variability in system operation, including variability in actual taxi times. In both HITL simulations, the controller was required to implement the "advice" of the automation. The initial and second HITLs involved two and six controller participants, respectively. Both simulations manipulated traffic level, but did not include off-nominal events. Hoang, Jung, Holbrook and Malik (2011) found that the spot release-time recommendations decreased controller situation awareness. Verbal reports from the controllers indicated that attending to the automation interfered with their own planning. Underlying the observed problems was a mismatch between the goals of the automated tool and the goals of controllers. The effects on situation awareness of a more recent version of the advisor that includes electronic flight data are still being assessed (Hayashi, 2012).

A European tool, Airport Collaborative Decision Making, aims to reduce gate turnaround time by coordinating the actions of airlines, airports, air traffic control and pilots. Each aircraft has a predetermined sequence and knowledge of the taxi time of a given flight. Pushback times can be given to every flight so that it can travel unimpeded to the runway and immediately take off. Successful implementation would require adoption of the automation by a large number of airports, and on accurate data supplied by all parties.

In the U.S. there is also a research effort focused on the enhancement of continuous, real-time collaboration between ramp and air traffic controllers (Fernandes, Smith, Spencer, Wiley & Johnson, 2011). Note that this collaboration adds a task to the traffic management coordinator's (TMC) repertoire. The Collaborative Airport Traffic System (CATS) is the current test-bed used to examine this concept, which has been successfully used in operations at JFK airport.

Capability 2: Taxi Routing and Scheduling

Several laboratories in the U.S. and Europe are testing systems that provide automatic generation of taxi routes (Cheng and Foyle, 2002; Stelzer; Morgan, McGarry, Klein and Kerns, 2011; Simaiakis et al., 2011) based on route data from all aircraft in the system, and data from real-time surface surveillance.

Testing of the Ground-operation Situation Awareness and Flow Efficiency tool (Go-SAFE) system (Verma et al., 2010) continues to elicit functional and interface problems (e.g. sub-optimal screen size and resolution, data-tag clutter issues, inappropriate color usage). Automated generation and delivery of clearances failed to lower controller workload, in part because of a mismatch between the clearance plans generated by automation and by controllers. Automation tended to generate unusual ad hoc taxi-route assignments while controllers preferred consistent familiar assignments. This is an example of a much more general clash between ad hoc flexibility, typically championed by operations research algorithms, and simplicity and consistency, typically championed by human controllers.

The Go-SAFE software did not allow a sufficiently flexible partnership between the controller and automation. As events unfolded, the automated agent often issued revised routes to the aircraft without controller agreement, and provided inadequate notification of the changes to controllers. Also the software did not permit the controller to modify clearances already issued, or to issue conditional clearances.

Mitre CAASD also has an airport surface automation testbed (Klein, Stelzer, Nelson, Brinton and Lent, 2010). HITL simulation results showed that controllers trained to have the same goals as the software always chose to follow the taxi routes suggested by software. The reluctance of controllers to issue route modifications was originally due to an unfriendly keyboard interface, but in a later study that added a map-based method for altering routes, controllers remained reluctant to amend automated taxi routes.

Capability 3: Departure Runway Assignment

No published HITL simulations were found for the Departure Runway Assignment capability, although preliminary discussions of the concept have been reported (Morgan, 2010).

Subject Matter Expert (SME) knowledge elicitation performed by NASA revealed that traffic management coordinators may consider up to 26 pieces of information to make a departure runway assignment. Because it is difficult for any human to effectively consider all these factors, particularly under periods of high workload, there is a clear need for automation designers to provide decision support tools. The OIs describe tools that suggest optimal runway assignments to the controller who can then accept or modify the assignment. However, automation designers must guard against the temptation to overweight factors that might help the operations-research goal of increasing capacity while underweighting the concerns of controllers. Where possible automation should be built to collect and integrate information that controllers would assemble manually. It is also important that there be a clear delineation between factors that automation has and has not factored into its recommendations so controllers know when omitted factors justify modifying automated advice.

Runway balancing is a function performed by the TMC, although individual runway assignments can be changed by controllers based on the immediate tactical situation. Atkins and Walton (2002) studied how well departure runways were currently balanced. They reported that traffic managers do not currently have accurate information about the future departure demand, nor the ability to predict how the surface situation will evolve, both of which are needed for effective TMC and controller decision-making.

Capability 4: Runway Scheduling

Jung et al. (2010) investigated an automated runway scheduler used in conjunction with the spot-release planner mentioned earlier. Together they have the potential for impressive reductions in the number and duration of stops in the queue. Surprisingly, however, the expected improvements in human performance (workload and situation awareness) did not materialize in practice (Hoang et al., 2011). Local controller workload and situation awareness were unchanged whether the runway scheduler was present or absent. However, controllers reported that they were mentally performing the runway scheduling task even when the advisor was present.

Capability 5: Departure Flow Management

The goal of the Departure Flow Management (DFM) tool is to use automation to improve the present manual process for releasing takeoffs, which requires a tower controller to make a phone call to Air Route Traffic Control Center (ARTCC). In the field trial, the tool was used in shadow-mode (Spencer, Carniol, Pepper & Smith, 2009). Controllers reported benefitting from two features of the tool: a record of what was done, and the seamless integration of new actions into the current operational picture. Several interface issues were identified: font size was too small to read the screen while standing, and an auditory cue was needed to signal when new information had arrived. A possible advantage of the DFM tool is that it provides the means to automatically communicate traffic management restrictions to the Traffic Flow Data Manager (TFDM). This kind of automatic sharing of information across tools and across installations is an important potential benefit from improved NextGen tools.

Doble, Timmerman, Carniol, Klopfenstein, Tanino & Sud (2009) reported the results of a field trial of a DFM tool. Tower controllers judged the tool to be useful, easy to use, and provided good access to needed information. Controllers also reported that the tool actually opened up more of their time for managing other issues, an important "figure of merit" that is rarely achieved.

So far, the Integrated Departure Route Planning (IDRP) tool has only been assessed via Subject Matter Expert surveys (Masalonis et al., 2008) therefore no conclusions will be drawn here.

Capability 6: Integrated Arrival/Departure Scheduling

There were no published HITL simulations found for this capability.

Arrival/Departure Management Tool (A/DMT) is actually a set of tools that will be a critical part of the tower controller's workstation. A/DMT will integrate information from surveillance (stored in an easily accessible database) and other DSTs to characterize arrival/departure demand and surface and airspace constraints. The vision is that it will integrate traffic flow constraints provided by DFM.

Capability 7: Runway Configuration Management

There were no published HITL simulations found for this capability.

The Runway Configuration Management (RCM) problem is to determine which runways should be used for arrivals or for departures. NASA Langley is developing System Oriented Runway Management (SORM) tools. SORM is a composite of two subsystems: Runway Configuration Management (RCM) and the Combined Arrival/Departure Runway Scheduling (CADRS) which assigns flights to runways in real time, accomplishing goals such as runway balancing (Lohr, Brown, Stough, Atkins, Eisenhawer & Long, 2011). A very small capacity increase was seen in computational simulations. In addition, it appears that even these small capacity gains are achievable with more frequent dynamic runway changes likely to increase complexity and controller workload. It is unclear whether this capability should be included in the FAA's mid-term plans.

Discussion

A great deal of NextGen resources have been channeled toward the difficult job of algorithm development and the identification of the NAS benefits (e.g., increased capacity, increased efficiency, reduced environmental impact) that can be expected if the algorithms are deployed. The operational and human performance results and tool maturity show some successes, such as the departure flow management tool that automates the manual process of ATC tower phone calls to ARTCC to release take-offs. Unfortunately, there are numerous tools that require considerable progress before the NextGen vision can be realized. Four of the seven capabilities have reached a level of maturity where they have been tested with humans in the loop; Departure Metering, Taxi Routing and Scheduling, Runway Scheduling and Departure Flow Management.

For all capabilities current operations have been explored, the proposed concept has been detailed and its application discussed, parameters of interest have been identified, and operational benefits that are likely to be realized highlighted. Weaknesses in the capabilities development include:

- Algorithm heuristics were rarely developed from knowledge about the human user's mental model,
- Algorithms often did not properly incorporate uncertainty,

- Initial validation using computational models of the operational system typically did not include sufficient (or sometimes any) representation of the human operator as a sub-system,
- High fidelity simulations were conducted with tools of inadequate maturity,
- HITL simulations neglected to collect both objective and subjective human performance measures,
- HITL simulations disregarded basic experimental design principles, and
- As the capabilities mature, testing continues in isolation of other capabilities.

It is important that individual tools be reliable. When an unreliable tool is introduced into tower operations, there is an increased likelihood that it will not be used as intended. Controllers may use the tool's recommendations during nominal low-workload conditions (where little help was needed anyway), but turn them off under off-nominal high workload conditions (precisely those conditions for which it was assumed that help was most needed), because the tool's recommendations for complex situations could not be trusted. Alternatively controllers may use a poorly designed tool for an unintended purpose (such as gaining access to raw information rather than an action recommendation), a purpose for which some other aid would have been more cost-effective. And of course, if using the tool is more trouble than it is worth, controllers may place it under the console and not use it at all.

With the simultaneous introduction of multiple tools anticipated for NextGen, it becomes even more critical to ensure that tools have reached a high readiness level. If unforeseen problems arise, it will be difficult to pinpoint which of the multiple new systems introduced is the source of the problem. Furthermore, joint use of multiple new tools is likely to produce emergent problems due to unforeseen and unstudied interactions among the tools. Of course there is a continuum in the degree to which new tools will interact in usage, but in general there is huge overlap in the information used by different tools, and in the operational impact of tools on traffic. It is implausible to make a "default assumption" that multiple new tools developed independently will play well together. The typical practice of developing each tool in isolation postpones the issue of properly integrating multiple tools into the future. But integration cannot be indefinitely postponed if tools are ever to reach operational status. Research into integrated suites of new tools has begun in the last few years. A further ramp-up of integration research is needed, including more HITLs testing integrated suites of multiple tools, useful for accomplishing multiple task goals. In the long run, integration can become a positive strength of tool-development research, providing benefits greater than the sum of individual tool benefits.

Acknowledgements

This work was funded under Interagency Agreement #DTFAWA-09-X-80020 between NASA and the FAA. The FAA Human Factors Division coordinated the research requirement and its principal representative acquired, funded, and technically coordinated execution of this research.

References

Atkins, S., Brinton, C., Walton, D., Arkind, K., Moertl, P. & Carniol, T. (2003). Results from the initial Surface Management System field tests. In *Fifth Eurocontrol/FAA Conference*, Budapest, Hungary.

Atkins, S. & Walton, D. (2002). Prediction and control of departure runway balancing at Dallas/Fort Worth airport. In *American Control Conference*, Anchorage, AK.

Beard, B. L. (2012). Human-in-the-loop research supporting mid-term NextGen air traffic control tower operational improvements. Downloadable at: *https://www.hf.faa.gov/hfportalnew/Search/SearchReport.aspx.*

Beard, B. L., Galeon, M. & Parke, B. (2013). Are NextGen enablers ready for prime time? Manuscript in preparation.

Beard, B. L., Parke, B., Holbrook, J., & Oyung, R. (2013). Survey of human factors experts on the potential human performance risks and benefits of NextGen capabilities. Manuscript in preparation.

Brinton, C. & Lent, S. (2012). Departure queue management in the presence of traffic management initiatives. Paper presented at the *Integrated Communications, Navigation and Surveillance Conference*, .

Cheng, V. H. L. & Foyle, D. C. (2002). Automation tools for enhancing ground-operation situation awareness and flow efficiency. *Proceedings of the AIAA Guidance, Navigation, and Control Conference, Paper AIAA 2002-4856*, Monterey, CA.

Doble, N. A., Timmerman, J., Carniol, T., Klopfenstein, M., Tanino, M., & Sud, V. (2009). Linking traffic management to the airport surface. *Proceedings of the Eighth USA/Europe Air Traffic Management Research and Development Seminar*, Napa, CA.

Fernandes, A. B., Smith, P. J., Spencer, A., Wiley, E., & Johnson, D. (2011). Collaborative Airport Traffic System (CATS) to evaluate design requirements for an airport surface departure management system. *Proceedings of* *the Human Factors and Ergonomics Society, 55*: 1827.

Government Accounting Office. (2010). NextGen Air Transportation System: FAA's Metrics Can Be Used to Report on Status of Individual Programs, but Not of Overall NextGen Implementation or Outcomes. GAO-10-629.

Hoang, T., Jung, Y., Holbrook, J. B., & Malik, W. A. (2011). Tower controllers' assessment of the spot and runway departure advisor (SARDA) concept. *Proceedings of the Ninth USA/Europe Air Traffic Management Research and Development Seminar*, Berlin, Germany.

Holbrook, J., Parke, B., Oyung, R., Collins, R., Gonter, K. & Beard, B. L. (2013). Perceived usefulness of planned NextGen tools by air traffic control tower controllers. *Proceedings of the 17th International Symposium on Aviation Psychology.* Dayton, OH: The Wright State University.

Holbrook, J., Puentes, A., Stasio, N., Jobe, K., McDonnell, L., & Beard, B. L. (2011). How thoroughly do proposed NextGen mid-term operational improvements address existing threats? *Proceedings of the 16th International Symposium on Aviation Psychology.* Dayton, OH: The Wright State University.

Hayashi, M. (2012). SARDA HITL preliminary human factors measures and analysis. Retrieved on February 27, 2012 from http://ntrs.nasa.gov/search.jsp?R=20120015369.

Jung, Y., Hoang, T., Montoya, J., Gupta, G., Malik, W., & Tobias, L. (2010). A concept and implementation of optimized operations of airport surface traffic. *Proceedings of the 10th AIAA Aviation Technology, Integration, and Operations Conference (ATIO)*. AIAA, Fort Worth, Texas.

Klein, K. A., Stelzer, E. K., Nelson, E. T., Brinton, C. & Lent, S. (2010). Improving efficiency with surface trajectory-based operations and conformance monitoring. *Proceedings of the 29th Digital Avionics Systems Conference*, Salt Lake City, UT.

Lohr, G. W., Brown, S., Stough, H. P. I., Atkins, S., Eisenhawer, S. & Long, D. (2011). System oriented runway management: A research update. *Proceedings of the Ninth USA/Europe Air Traffic Management Research and Development Seminar*, Berlin, Germany.

Masalonis, A., Bateman, H. DeLaura, R., Song, L. Taber, C. & Wanke, C. (2008). Integrated departure route planning. *Proceedings of the 27th Digital Avionics System Conference,* St. Paul, MN.

Monroe, G. A. (2009). Surface Management System departure event analysis. NASA/TM-2009-215383.

Morgan, C. E. (2010). Mid-term Surface Trajectory-Based Operations (STBO) Concepts of Use: Departure Runway Assignment, MTR100269V4.

NextGen Implementation Plan (2012). Retrieved February 27, 2012 from *www.faa.gov/nextgen/implementation/plan/*.

Simaiakis, I., Khadilkar, H., Balakrishnan, H., Reynolds, T. G., Hansman, R. J., Reilly, B., & Urlass, S. (2011). Demonstration of reduced airport congestion through pushback rate control. *Proceedings of the Ninth USA/Europe Air Traffic Management Research and Development Seminar*, Berlin, Germany.

Simaiakis, I., Sandberg, M., Balakrishnan, H. & Hansman, R. J. (2012). Design, Testing and Evaluation of a Pushback Rate Control Strategy. *Proceedings of the Fifth International Conference on Research in Air Transportation*, Berkeley, CA.

Spencer, A., Carniol, T., Pepper, J. & Smith, P.J. (2009). Airport departure flow management (DFM): Findings from field trial testing. *Proceedings of the 15th International Symposium on Aviation Psychology*, Dayton OH.

Stelzer, E. K., Morgan, C. E., McGarry, K. A., Klein, K. A. & Kerns, K. (2011). Human-in-the-loop simulations of surface trajectory-based operations. *Proceedings of the Ninth USA/Europe Air Traffic Management Research and Development Seminar*, Berlin, Germany.

Verma, S., Kozon, T., Lozito, S., Martin, L., Ballinger, D., & Cheng, V. (2010). Human factors of precision taxiing under two levels of automation. *Air Traffic Control Quarterly, 18(2),* 113-141.

ASSESSING THE CHANGING HUMAN PERFORMANCE RISK PROFILE IN THE NEXTGEN MID-TERM

Katherine A. Berry & Michael W. Sawyer
Fort Hill Group
Washington, DC

Many Next Generation Air Transportation System (NextGen) Operational Improvements (OIs) aim to provide controllers with decision support tools and other automation specifically designed to provide safety enhancements to National Airspace System (NAS) operations. While these changes may indeed produce positive safety improvements, the introduction of each new system and capability also offers the possibility of introducing new human performance hazards into the NAS. A thorough review of the proposed NextGen midterm OIs was completed to identify the potential for both the positive and negative impacts on the human contribution to risk in the NAS. A summary of these findings was presented by linking the proactive human performance assessment with the Federal Aviation Administration's (FAA) top five hazards in the NAS for fiscal year 2013. The results showed that while some of the human performance hazards present in current operations were reduced or eliminated, many new human performance hazards could also be introduced as these systems are implemented.

The FAA is currently executing a considerable transformation of the NAS called NextGen, which aims to improve the convenience and dependability of air travel while increasing safety and reducing environmental impact. NextGen plans to meet these goals by introducing a variety of new systems and capabilities, such as the introduction of data communications and Automatic Dependent Surveillance-Broadcast (FAA, 2012). NextGen ventures to improve the capacity, efficiency, and, perhaps more importantly, safety of the NAS through the implementation of OIs. Many operational improvements aim to provide controllers with decision support tools and automation specifically designed to provide safety enhancements to NAS operations. While NextGen may produce many positive safety improvements, the introduction of each new system and capability also offers the possibility of increasing the human contribution to risk in the NAS (Sawyer, Berry, & Blanding, 2011). This is especially true when considering the system-wide impact and concurrent development of many of the systems (Berry & Sawyer, 2012). From a risk management perspective, research into these effects is needed to address the potential for both positive and negative impacts on the safety of the NAS (FAA, 2011).

The human factors community of practice has played a significant role in the enhancement of safety in the aviation and air traffic control (ATC) domains. However, in terms of traditional safety assessments, many of the tools and techniques utilized by human factors practitioners are often retrospective in nature. These techniques and tools serve as aids in the analysis of incident and accident data gathered post-hoc. While these tools and techniques have been and are still valuable in assessing the safety of the NAS, these tools and techniques are limited in regards to analysis of future systems, such as NextGen (GAO, 2011; FAA, 2011). The ability to proactively identify potential hazards to human performance associated with a new system before the new system is introduced into the NAS has long been identified as a need (GAO, 2010; GAO, 2011). Since NextGen is introducing several new systems and capabilities, the integration of proactive human factors safety research into the earliest stages of system design and acquisition could not only reduce industry cost, but also improve system design, development, and implementation (EUROCONTROL & FAA, 2010). A standardized approach to proactively identifying and assessing human performance hazards is needed to ensure these hazards are identified, described, and tracked. The purpose of this proactive safety assessment is to examine the potential changes in the human performance risk profile of current NAS operations with NAS operations in specified NextGen timeframe. Furthermore, the paper will present a summary of results of NextGen human performance assessments (Berry, Sawyer, & Austrian, 2012; Berry & Sawyer, 2012) in comparison with FAA Air Traffic Organization's (ATO) list of top hazards in the NAS for FY2013.

Methodology

This analysis will focus on the Segment Bravo portion of the NextGen midterm timeframe. The midterm and Segment Bravo are comprised of seven solution sets represents changes planned to be implemented beginning in 2015 (DOT IG, 2012). The solution sets are a compellation of related operational improvements grouped by overarching themes (FAA, 2013). Each solution set contains multiple OIs with each OI proposing a new system or capability. The comparisons in study will be made at the OI level.

The Human Error and Safety Risk Analysis (HESRA) tool was utilized to perform the assessment of the NextGen midterm Segment Bravo OIs (FAA, 2009). Based on a traditional Failure Modes and Effects Analysis, the HESRA tool provides a structured method for identifying potential human performance hazards and the associated worst credible outcome. A panel of ATC experts and human factors experts also estimated the severity, likelihood, and recovery potential for each potential human performance hazard. The FAA's Safety Management System likelihood and severity scales (FAA, 2008) were utilized in order to ensure that identified hazards could be managed in the existing FAA safety process. The recovery rating in the HESRA methodology assesses the ability to both detect and recover from the presence of hazards in the system. Recovery is essential to differentiate between hazards that will be detected and recovered from before the effects propagate through the system and hazards that will have major system effects before they are identified and corrected. These values are then combined to calculate a Risk Priority Number, which can be used to compare the relative risk associated with identified potential human performance hazard.

The FAA's ATO releases an annual listing of top hazards in the NAS each fiscal year (FY). The list is generated from extensive reviews of safety data and discussions with ATO stakeholders. The list, which represents the top safety hazards, aims to assist in prioritizing ATO resources in improving the safety of the NAS. The top five hazards list for FY2013 is provided in Table 1 (Teixeira, 2013). For this assessment, the FY2013 top five hazards were compared to the Segment Bravo assessment results in an effort to determine linkages between the two analyses.

Table 1. FY13 Top Five Hazards (Teixeira, 2013)

FY 2013 Top 5 Hazards
Recovery In some cases, margins of safety are not quickly re-established after a loss of separation.
Traffic Advisories/Safety Alerts Safety alerts and/or traffic advisories are not being issued when required, removing a safety barrier and increasing risk.
Failure to Monitor Initial Departure Headings Communications are being transferred prior to ensuring initial departure headings, resulting in aircraft being off frequency while controllers attempt to mitigate losses of separation.
Similar Sounding Call Signs Aircraft are operating with similar sounding calls signs, resulting in increased opportunities for confusion, and incorrect aircraft receiving or reading back clearances.
Conflicting Procedures Facility letters of agreement and standard operating procedures conflict with published arrival and/or departure procedures, increasing the likelihood for incorrect pilot readback and actions.

Results and Discussion

Sample potential human performance hazards from the Segment Bravo human performance hazard assessment were identified and linked to the FY2013 top five hazards. In the following sections, the linkage to the top five will be discussed, and additional human performance risk hazards will be identified.

FY2013 Top Five Hazards Linkage

Table 2 outlines sample potential human performance hazards that are associated with one of the ATO's top five hazards from FY2013.

Table 2. FY13 Top Five Hazards Linkage

Top 5 Hazard	OI	Hazard Condition	Potential Human Performance Hazard	Worst Credible Outcome
Recovery	102144	Expanded use of Automated Terminal Proximity Alert, which alerts approach controllers regarding separation problems for in-trail aircraft on final approach, to include use during complex dependent approach operations involving mixed wake categories.	Controller over-relies on automated terminal proximity alerting system and does not actively monitor spacing. Automation fails to display alert when necessary.	Controller fails to notice impending loss of separation. Potential for wake turbulence encounter.
Alerts	102406	Surface surveillance equipment is compatible with runway incursion indicating and alerting capabilities, so that controller and pilots would be warned if a ground vehicle enters an active runway, in a manner similar to that for an intruding aircraft.	Controller fails to notice alert regarding unauthorized ground vehicle entering an active runway due to alert salience.	Ground vehicle enters active runway. Potential for runway incursion between aircraft and ground vehicle.
Heading	102114	Monitor aircraft conformance and provide an alert on the Radar Console when an aircraft's track or position information indicates that the aircraft is laterally deviating from its assigned route.	Controller fails to notice alert regarding aircraft laterally deviating from its assigned route.	Aircraft continue on present trajectories until controller resolves issue or short term conflict alert activates.
Call Signs	104207	Initial D-TAXI service will be the principle communication of taxi-out instructions via data communications instead of voice, from ground controllers in the ATC tower to flight crews as their departing flights enter the movement area of the airport surface and head to their assigned runway.	Controller sends taxi instruction to incorrect aircraft due to mis-read of call signs. Taxi route is inadequate for aircraft characteristics.	Aircraft taxis via inadequate taxiway instruction. Potential for taxiway incursion. Controller re-routes aircraft to correct destination. Potential for ground delays.

Top 5 Hazard	OI	Hazard Condition	Potential Human Performance Hazard	Worst Credible Outcome
Conflicting Procedures	102114	Transfer of radar identification will be fully automatic and accomplished by the automation without any controller activity if the aircraft is problem free at the transfer point.	Transfer point identified by automatic handoff automation is inadequate. Transfer point is not in agreement with prior letter of agreement.	Aircraft automatically hands-off too early or too late to be in accordance with letter of agreement. Transferring controller inhibits automatic handoff automation and manually coordinated handoffs with receiving controller.

Recovery. The top five hazard of recovery is related to the ability of a controller to issue the necessary instructions to safely recover from an adverse safety event, such as a loss of separation or airspace violation. From a human performance perspective, this action is composed of two tasks: detecting the presence of an adverse event, and executing a strategy to reestablish positive safety. Since NextGen introduces many new decision support tools and additional automation, the potential exist for controller's to suffer skill degradation due to a reliance on automation. In the sample human performance hazard, automation may support the controller by monitoring the spacing of aircraft during final approach. However, the hazard exist that the automation may not be perfect and in a particular instance, many not identify a spacing issue. If the controller has become over-reliant on the automation over time, the controller may also not identify the impending conflict. While this hazard might not occur frequently, the recovery period and ability for the controller has been reduced due to over-reliance on automation and skill degradation. This potential human performance hazard and other similar hazards should be mitigated by training requirements that provide recovery training to controllers by improving the controller's ability to detect a problem occurring and to develop and execute a plan to correct the problem at hand.

Traffic Advisories/Safety Alerts. The top five hazard of traffic advisory/safety alert centers around the problems associated with safety alerts not being issued by controllers to pilots when the operational conditions present necessitate an alarm. In many situations, controllers have more knowledge regarding the conditions of the airspace or airport. For example, controllers may have more detailed weather information or traffic information. Relaying the information to pilots is necessary for ensuring the safety of the NAS. Currently, controllers proactively issue traffic advisories, and NextGen aims to assist the controller by providing decision-support tools to assist controllers in identifying when to issue such advisories. In the sample human performance hazards, ground automation may monitor ground vehicle movement in conjunction with aircraft movement. If a ground vehicle enters or is on a trajectory appearing to enter an active runway, automation will provide the controller with an alert regarding the vehicle. If the controller does not notice the alert because the alert is not salient enough for the environment, the controller may remain unaware of the ground vehicle and not issue the necessary instructions and advisories to avoid a runway incursion. This potential human performance hazard and similar hazards should be mitigated by design and research requirements that ensure the automation alert assisting the controller in the issuance of traffic advisories is salient and convey the correct information in the appropriate amount of time.

Failure to Monitor Initial Departure Headings. The top five hazard of failure to monitor initial departure headings is related to the controller's task of monitoring the aircraft's adherence to a specific route and, in particular, the initial departure heading. With NextGen, the usage of area navigation (RNAV) routes will be increased in an effort to reduce communications and improve efficiency. RNAV routes in conjunction with NextGen monitoring tools aim to reduce track deviations. In the sample human performance hazard, automation may monitor aircraft conformance to an assigned route. If the aircraft laterally deviates from the route, the automation may alert the controller. It is possible that a controller may fail to notice the track deviation alert resulting in a track deviation and potential conflict. While a decision support tool may assist in identifying potential track deviations, the possibility always exists for the automation to be imperfect or for the controller to not perceive the automation alert. Like the previous hazard, the salience of the automation notification or alert should be examined. However, other mitigation strategies exist for this particular top five hazard. For example, prior to departure a controller can verify the initial departure heading or first departure fix with the pilot. A procedural mitigation strategy may be implemented to address the hazard during the development of the automation strategy and may support the automation after deployment.

Similar Sounding Call Signs. The top five hazard of similar sounding call signs identifies the issue of confusion associated with call signs and hearback/readback errors. With NextGen, the usage of voice communications may be supplemented with the increase usage of data communications. While issues particular to voice communications, such as similar sounding call signs, may be reduced, new human factors issues particular to data communications may arise (e.g., misreading similar looking call signs). In the sample human performance hazard, the ground controller to may utilize data communications to issue pilots taxi instructions. If the controller mis-reads a call sign, the potential exist for the controller to send the incorrect taxi instruction to the mis-read aircraft. The aircraft could then taxi via an inadequate taxi route. Future research should be conducted to further examine the mis-read of data communications for both the controller and pilot.

Conflicting Procedures. The tops five hazard of conflicting procedures is related to operational procedures not being in accordance with existing procedures or letter of agreements. NextGen operational improvements may implement many new procedures and systems that must interact with existing procedures and letter of agreements. It is important for designers to be familiar and incorporate those existing conditions and procedures into the NextGen capabilities. In the sample human performance hazard, the handoff of an aircraft's radar identification to the bordering sector may be automated to reduce controller workload. Once a aircraft reaches a specific transfer point in a sector, the aircraft will automatically handoff to the receiving sector. However, if designers have not incorporated an existing letter of agreement in the determination of the transfer point, the transfer point could be incorrectly placed resulting in an aircraft handing-off at the in proper time or manner. If the problem of the incorrect transfer point continues, the transferring controller may elect to inhibit the automatic handoff feature eliminating the benefits associated by the automation. This human performance hazard and similar hazards should be mitigated by developing design requirements for the incorporation of existing procedures and letter of agreements into new systems.

NextGen Human Performance Assessment

The ATO's top five hazards are associated with many of the NextGen human performance assessment hazards. Furthermore, the NextGen human performance assessment identified many other potential risks. For more detail information on portions of the NextGen human performance assessment please see Berry and Sawyer (2012), Berry, Sawyer, and Austrian (2012), and Sawyer, Berry, and Austrian (2012). These assessments and the overall assessment identified many human performance hazards associated with Segment Bravo, and also developed design, training, and research requirements targeted to mitigate or eliminate the impact of those hazards. The overall findings included human performance issues associated with data communications on the ground, enhanced vision systems for pilots, integrated en route display systems, inter-relationships of automated systems, and proposed alerts and notifications.

To expand on one of the human performance findings identified by the overall assessment, the findings presented by Sawyer, Berry, and Austrian (2012) revealed that many new alerts and notifications could potentially be implemented into the NAS in the NextGen midterm. Whereas current air traffic controllers have two primary safety-critical alerts, the conflict alert and minimum safe altitude warning, controllers in the midterm could have an additional nine alerts or notifications related to the implementation of new procedures with reduced separation standards. As each of these alerts is developed and implemented, a series of research studies will be needed to determine the proper implementation strategy and to determine if the potential for over-burdening the controller with too many alerts and too much information exist. These efforts should focus on determining the necessary mode, salience, location, required level of accuracy, and place in the greater air traffic alert hierarchy for each alert.

Conclusions

In an effort to improve the efficiency and safety of the NAS, NextGen introduces new automation and capabilities in the form of operational improvements. While these changes may indeed produce positive safety improvements, the introduction of each new system and capability also offers the possibility of introducing new human performance hazards into the NAS. A thorough review of the proposed NextGen midterm Segment Bravo was completed to identify the potential for both the positive and negative impacts on the human contribution to risk in the NAS. A summary of these findings was presented by linking the proactive human performance assessment with the ATO's top five hazards in the NAS for fiscal year 2013. The results showed that while some of the human performance hazards present in current operations were reduced or eliminated, many new human performance hazards could also be introduced as these systems are implemented.

Acknowledgements

We would like to acknowledge the FAA's Human Factors Division (ANG-C1) for funding this project and similar work. Additionally, we would like to acknowledge the air traffic control and human factors subject matter experts who help to assess the operational improvements.

References

Berry, K. & Sawyer, M. (2012). Assessing the Impact of NextGen Trajectory Based Operations on Human Performance. In the *Proceedings of the 4th Annual Applied Human Factors and Ergonomics Conference*, 2012, San Francisco, CA.

Berry, K., Sawyer, M., & Austrian, E. (2012). Behind the scenes of NextGen: describing the impact of NextGen operational improvements on the traffic manager. In the *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 2012, Boston, MA.

Department of Transportation Inspector General. (2012). *Status of Transformational Programs and Risks to Achieving NextGen Goals* (AV-2012-094). Retrieved from http://www.oig.dot.gov/sites/dot/files/NextGen%20Transformational%20Programs%5E4-23-12.pdf.

EUROCONTROL & Federal Aviation Administration. (2010). *Human Performance in Air Traffic Management Safety*. EUROCONTROL/FAA Action Plan 15 Safety

FAA. (2008). Air Traffic Organization: Safety Management System Manual. Retrieved January 2012, from http://www.faa.gov/about/safety_efficiency/sms/resource.

FAA. (2009). Human Error and Safety Risk Analysis for Federal Aviation Administration Air Traffic Control Maintenance and Operations. Retrieved 2012, from http://www.hf.faa.gov

FAA (2011). *Destination 2025*. Retrieve 2011, from www.faa.gov/about/plans_reports/media/Destination2025.pdf.

FAA. (2012). NextGen Implementation Plan. Retrieved March 2012, from http://www.faa.gov/nextgen

FAA. (2013). NAS Enterprise Architecture. Retrieved February 2013, from https://nasea.faa.gov/

Government Accountability Office. (2010). *Aviation Safety: Improved Data Quality and Analysis Capabilities Are Needed as FAA Plans a Risk-Based Approach to Safety Oversight* (GAO-10-414). Retrieved from http://www.gao.gov/products/GAO-10-414.

Government Accountability Office. (2011). *Aviation Safety: Enhanced Oversight and Improved Availability of Risk-Based Data Could Further Improve Safety* (GAO-12-24). Retrieved from http://www.gao.gov/products/GAO-12-24.

Sawyer, M., Berry, K, & Blanding, R. (2011). Assessing the Human Contribution to Risk in NextGen. In the *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2011, Las Vegas, NV.

Sawyer, M., Berry, K., & Austrian, E. (2012). Analysis of new proposed air traffic control alerts in the NextGen midterm. In the *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 2012, Boston, MA.

Teixeira, J. (2013). *Improving Aviation Safety: An Air Traffic Control Perspective* [PowerPoint Slides]. Retrieved from https://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/safety/

ENHANCING CREW RESOURCE MANAGEMENT TRAINING PROGRAM:
THE INTRODUCTION OF A COGNITIVE-ADAPTATION TRAINING

Marie-Pierre Fornette
Institut de Recherche Biomédicale des Armées
Brétigny, France

Marie-Héloïse Bardel, Camille Lefrançois, Jacques Fradin
Institut de Médecine Environnementale
Paris, France

Farid El Massioui
Université Paris 8
Saint-Denis, France

Françoise Darses, Corinne Roumes
Institut de Recherche Biomédicale des Armées
Brétigny, France

The goal of "crew resource management" (CRM) training programs is to enhance safety and efficiency in flight. To achieve this goal, CRM training have to teach the appropriate knowledge and skills, and thus, to be adjusted according to all environmental changes. During the last decade, there has been an increasing need to reinforce skills of French military crews, to deal with complex and unforeseen situations. A new cognitive-adaptation training was proposed. It specifically seeks to enhance metacognitive skills related to the two main types of human cognitive processes, and to strengthen reflective processes involved in cognitive and emotional adaptation. The effects of this training on adaptation skills and the relevance for the CRM courses of theoretical adaptation knowledge pertaining to this new training were evaluated. The results indicate that the new training may provide a useful tool for enhancing the ability of flying crews to manage complex and unforeseen situations.

To achieve high levels of safety and performance during flights, pilots need both technical and non-technical skills. For more than three decades, human factor (HF) training programs known as "crew resource management" (CRM) training have been used to teach non-technical skills to pilots. The goal of CRM training is to improve safety and efficiency in flight. The literature and aviation regulations provide recommendations on a set of core topics (e.g., communication, workload management, decision-making) which must be taught in CRM training. However, there is a great variability in what is trained during a CRM course (Salas, Wilson, Burke, & Wightman, 2006). In France, military-aviation CRM courses teach pilots general HF topics (e.g., communication, situation awareness, decision-making, fatigue), as well as more specific topics (e.g., different types of violations, military deployment, systemic safety). This is done through interactive training methods involving discussions regarding practical experiences by the participants.

To be effective, CRM training programs must be adjusted over time to remain relevant in a rapidly changing operational, organizational, and technical environment. New CRM training needs can be identified, on the one hand, through analyses of successful in-flight performance, and on the other hand, through the analysis of accidents, incidents or dysfunctions (O'Connor, Flin, & Fletcher, 2002). Such analyses can pinpoint CRM skills that need improvement through training. The topics included in French military-aviation CRM courses are based on an evaluation of needs which relies on, (a) observations and interviews of flying crews concerning their activities and HF difficulties, for the design of the CRM training, (b) evaluations of CRM training programs by the flying crews, after the training, and (c) analyses of changes in work conditions, over time. During the last decade, the French Air Force (FAF) was faced with an increasing number of constraints and changes due to the ever-increasing complexity and diversity of military operations and systems. In this context, operational task demands are likely to exceed the adaptation capabilities of military flight crews, and therefore, they are also likely to generate stress, to reduce performance, and to endanger air safety. An analysis of reports from the French Defense Air Accident Investigation Board reveals that some pilots were unable to use adequate adaptation skills to deal successfully with complex and unforeseen situations. Therefore, it seems especially important to enhance FAF CRM training programs with specific knowledge and tools targeting adaptation to, and management of, complex and unforeseen situations. Accordingly, a new training that seeks to improve cognitive- and emotional-adaptation skills was proposed to the FAF. The new training was evaluated in a first study, which involved a sample of pilot cadets. In addition, for more than a year, theoretical knowledge pertaining to this new training has been included into CRM courses for the FAF.

In the following sections we, first, review relevant literature on cognitive and emotional adaptation, which provides the background for the new cognitive-adaptation training. We then describe the main principles of this new training, the main results that have been obtained in the first evaluation of the training in pilot cadets, and the recent introduction of adaptation knowledge into the CRM courses. Lastly, we discuss potential applications of the new training.

Cognitive and Emotional Adaptation

To maintain high levels of performance and safety in complex, unforeseen, and risky situations, pilots must be able to use adequate cognitive strategies and to manage their stress. HF experts have developed training programs that seek to enhance, on the one hand, cognitive adaptation, and on the other hand, stress management. Cognitive-adaptation training typically targets metacognitive skills, which involve a reflection on cognitive processes such as decision making or situation awareness. The most commonly used form of stress-management training in military aviation is based on the notion of *Stress Exposure Training* (SET). SET seeks to counter the effects of stress on performance through physiological, cognitive, or collective methods; however, it does not usually address the causes of stress.

To address the need to improve training in the management of complex and unforeseen situations, we decided to benefit from recent findings in neuropsychology and neuroscience which shed light on adaptation. In particular, we investigated research focusing on the relationships between cognition and emotion.

Cognitive Adaptation

Cognitive and behavioral adaptation in complex and dynamic situations is based on cognitive control, which involves the dynamic adjustment of cognitive processes to situational features such as environmental demands, as well as human features (e.g., aircraft failure, pilot's perceived workload; Hoc & Amalberti, 2007). Continuous cognitive control allows pilots to adjust performance to changing situations, so as to maintain adequate efficiency and avoid excessive fatigue. In cognitive psychology, human cognition is usually modeled as a juxtaposition of two complementary types of cognitive processes (for review, see Evans & Frankish, 2009). The first type of cognitive processes, which are referred to as *automatic processes*, is usually involved primarily in the management of simple and well-known situations by using learned action strategies or innately programmed behaviors. The second type of cognitive processes, which are also referred to as *reflective processes,* relates to abstract reasoning, hypothetical thinking, and cognitive flexibility, which are involved in intervening on, or overriding of, default, automatic processes. Reflective processes play an important role in the management of complex and unknown situations via the formation of new, adaptive, and creative cognitive strategies, and they allow the individual to anticipate future events.

Emotion Regulation

In military operations, stress is obviously an important issue, which is related to the broader topic of emotion regulation. Three types of intentional emotion-regulation strategies are described in the literature: (a) avoidance of emotionally meaningful stimuli, (b) modulation of emotional responses, and (c) cognitive change or reappraisal, which involves changing how we think about a situation (Gross, 2002). Cognitive change appears to be one of the most effective strategies (Butler, Chapman, Forman, & Beck, 2006; Gross & John, 2003), presumably because it addresses directly the cause of emotions. It is largely used in cognitive-behavioral therapy through metacognitive techniques such as cognitive restructuring, which seek to improve access to more adaptive modes of thinking, or representations of events that trigger stress.

Relationships between Cognitive and Emotional Adaptation

In recent years, several studies have investigated the relationship between cognitive processes and emotion regulation (e.g, Compton et al., 2011; Ochsner & Gross, 2005). Functional imaging studies have revealed that cognitive change techniques can reduce emotional disturbance by inhibiting automatic bottom-up processes while strengthening reflective top-down processes in patients as well as in healthy individuals (e.g., Clark & Beck, 2010). Therefore, the awareness of, and self-adjustment to, one's own cognitive processes and representations appear to occupy a central place in both cognitive and emotional adaptation.

Neuropsychological studies focusing on adaptation disorders are consistent with this view. The neuropsychological concept of executive functioning shares several aspects with reflective processes. Executive functioning is described as: (a) facilitating adaptation when the novelty or complexity of a given situation precludes an automatic response, and (b) involving in emotion regulation by relying on higher-level cognitive processes (Suchy, 2009). Reflective (or executive-functioning) processes are likely essential for efficient self-regulation of cognition and emotion in complex or unforeseen situations.

A New Cognitive-Adaptation Training

Accordingly, we decided to propose a new training that seeks to enhance metacognitive skills related to the two previously mentioned types of human cognitive processes, and to strengthen reflective processes implied in cognitive and emotional adaptation.

Principles of the New Cognitive-Adaptation Training

This cognitive-adaptation training called *Mental Mode Management* (Fradin, Aalberse, Gaspar, Lefrançois, & Le Moullec, 2008; Fradin, Lefrançois, & El Massioui, 2006) is based on the Cognitive-Processes Scale (CPS). This scale provides a description of automatic processes (hereafter referred to as *automatic mental mode*), and of reflective processes (hereafter referred to as *adaptive mental mode*). Each mental mode is described using six dimensions (see Table 1). CPS is a pedagogic self-report tool, which consists of seven Likert-type scales, including six scales that are used to assess mental mode or frame of mind, and one scale used to assess stress. For example, the first scale allows individuals to evaluate whether they are approaching a situation as if it were routine (i.e., known and mastered), or with a curious frame of mind; the third scale is used to evaluate whether, in a given situation, the individual tends to think of "black and white", or in a more nuanced way. The training consists in:

1. choosing a stressful or difficult-to-adapt-to situation,
2. becoming (more) aware of the predominant frame of mind (or mental mode) and stress experience during that situation, using imagined confrontation and the CPS,
3. practicing (a) mental mode management (MMM) technique(s) to reinforce adaptive mental mode,
4. evaluating again the mental mode and stress experience in the chosen situation using CPS, to assess the (possible) difference pre-to-post practice and then the efficiency level of MMM technique(s).

MMM techniques are mainly reflexive. One of the most global techniques to promote the adaptive mental mode consists in asking oneself questions without providing immediate answers, for the six dimensions of this mental mode. Two example questions for the dimensions "logical reasoning" and "individual opinion" are: "How would it work if I looked at the situation in terms of causes and effects?" and "If I put aside the judgment of the others and I think about what is really at stake for me, what do I personally think? ", respectively (for more information about MMM techniques, see Fornette et al., 2012; Fradin, 2003; Fradin et al., 2008).

Table 1.

Dimensions of Automatic Mental Mode and of Adaptive Mental Mode.

Automatic Mental Mode	Adaptive Mental Mode
Routine	Curiosity
Refusal	Acceptance
Dichotomy	Nuance
Certainty	Relativity
Priority to results	Logical reasoning
Social image	Individual opinion

Effects of the New Training on Performance and Stress Management

A first evaluation of MMM training effects on flight performance and stress management was carried out in a sample of FAF pilot cadets. The main methodological features and results of this study may be summarized as follows (for additional information see: Fornette et al., 2012).

The class of pilot cadets ($N = 21$) was divided into two groups: a Training Group (TG), which participated in six two-hour training sessions, and a Control Group (CG), which did not receive training. Both groups were balanced with respect to emotional profiles, initial performance, and instruction-squadron membership. Within each group, cadets were further divided into two subgroups (High- and Low-performance level) based on the median score of the class. This resulted in four subgroups: TG-Low, TG-High, CG-Low, and CG-High. In-flight performance (in the form of scores ranging from 0 to 20, which were assigned by flight instructors) was measured; so were mood, anxiety, and stress-management mode, using questionnaires (POMS, STAI-Y-A, and specific questionnaires, respectively).

A comparison between in-flight performance before and after training showed a significant improvement ($p \leq .05$) for the lowest-ranked cadets in the training group (TG-Low). For the three other subgroups, no significant change was observed. Pre- and post-training flight score means of the four subgroups are shown in Table 2. The improvement for the TG-Low group persisted until the end of the basic flying program (i.e., 1.5 months after the end of the training). Mood and anxiety scores did not differ significantly between the training and control groups. However, the number of cadets who reported having changed their mode of stress management during the study was significantly higher ($p \leq .05$) for the training group (80%) than for the control group (27%). Moreover, 70% of the training group cadets stated that the cognitive-adaptation training had allowed them to better understand events and, consequently, to reduce their stress level.

Table 2.
Flight Score Means and Standard Deviations for the Low- and High-Level Subgroups of the Training and Control Groups During the Pre-training and Post-training Phases.

		Phase			
		Pre-training		Post-training	
Subgroup	*n*	*M*	*SD*	*M*	*SD*
TG-Low (Training Group - Low Level)	6	13.33	0.67	14.18	0.53
TG-High (Training Group - High Level)	4	14.03	1.01	13.65	0.67
CG-Low (Control Group - Low Level)	5	13.55	0.83	13.43	0.41
CG-High (Control Group - High Level)	6	14.34	0.66	14.33	0.59

Relevance of Theoretical Knowledge Pertaining to the New Training, for the CRM Courses

Knowledge concerning reflective processes and their involvement in cognitive and emotional adaptation has been inserted in introductory or refresher CRM courses in the FAF for over a year now. Such knowledge has been used to revise the “Stress management” topic of the CRM course, and to include a new topic called “Management of complexity and of the unforeseen”. The impacts of these revised and new topics were evaluated using four criteria and a four-point Likert scale (see Table 3). This is some of the usual evaluation of CRM topics which takes place just after the training. Although they are still preliminary, the results of this evaluation indicate that the inclusion into CRM courses of knowledge concerning cognitive and emotional adaptation enhances HF knowledge and provides flying crews with a new, richer, view of their own activities. The evaluation of the impact on the change of practices is also satisfactory, although, so far, only theoretical knowledge concerning adaptation has been introduced into CRM courses. This contribution to the change of practices should be further improved because MMM techniques will likely be included in future refresher CRM courses.

Table 3.
Percentage of responses "*Not at all*", "*Rather no*", "*Rather yes*", "*Definitely yes*" *for the four evaluation criteria corresponding to the CRM topics:* "*Stress management*" *and* "*Management of complexity and of the unforeseen*".

Criterion: Is the topic useful to…?	Not at all	Rather no	Rather yes	Definitely yes
1. becoming aware of the scope of FH dimensions	0.5 %	1 %	47 %	51.5 %
2. learning of new FH knowledge	0 %	0 %	27 %	73 %
3. promoting useful discussions	0.5 %	0.5 %	34 %	65 %
4. helping change practices	0.5 %	7 %	36.5 %	56 %

Conclusion

Although the evaluation carried out in pilot cadets remains limited, and the introduction in the CRM courses is relatively recent, the results suggest that the new cognitive-adaptation training has the potential to benefit cognitive and emotional adaptation skills. These results, if they are confirmed in future studies, would support the view that cognitive-adaptation training programs that seek to promote the use of reflective processes may provide a useful pedagogical tool for enhancing both cognitive and emotional adaptation in flying personnel, and ultimately, for improving flight safety.

The metacognitive skills acquired through this training on general adaptation skills can be generalized and transferred, which could be especially advantageous in risky and ever-changing occupational settings. Therefore, the principles of this cognitive-adaptation training could be used to supplement training programs that seek to improve situation awareness, decision making, stress management, and more generally, the management of complex and unforeseen situations, in various fields and for different operators' profiles (from novice to expert).

References

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review, 26*(1), 17-31. doi:10.1016/j.cpr.2005.07.003

Clark, D. A., & Beck, A. T. (2010). Cognitive theory and therapy of anxiety and depression: Convergence with neurobiological findings. *Trends in Cognitive Sciences, 14*(9), 418-424. doi:10.1016/j.tics.2010.06.007

Compton, R. J., Arnstein, D., Freedman, G., Dainer-Best, J., Liss, A., & Robinson, M. D. (2011). Neural and behavioral measures of error-related cognitive control predict daily coping with stress. *Emotion, 11*(2), 379-390. doi:10.1037/a0021776

Evans, J. S. B. T., & Frankish, K. (2009). *In two minds: Dual processes and beyond*. New York, NY: Oxford University Press.

Fornette, M.-P., Bardel, M.-H., Lefrançois, C., Fradin, J., El Massioui, F., & Amalberti, R. (2012). Cognitive-adaptation training for improving performance and stress management of airforce pilots. *The International Journal of Aviation Psychology, 22*(3), 203-223. doi:10.1080/10508414.2012.689208

Fradin, J. (2003). Gestion du stress et suivi nutritionnel. [Stress management and nutritional follow-up]. *Médecine et Nutrition, 39*(1), 29-34.

Fradin, J., Aalberse, M., Gaspar, L., Lefrançois, C., & Le Moullec, F. (2008). *L'intelligence du stress* [Intelligence of stress]. Paris, France: Eyrolles.

Fradin, J., Lefrançois, C., & El Massioui, F. (2006). Des Neurosciences à la Gestion du Stress devant l'Assiette. [Eating and managing stress with the help of neurocognitive therapy]. *Médecine et Nutrition, 42*(2), 75-81.

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology, 39*, 281-291. doi:10.1017.S0048577201393198

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*, 348-362. doi:10.1037/0022-3514.85.2.348

Hoc, J. M., & Amalberti, R. (2007). Cognitive control dynamics for reaching a satisficing performance in complex dynamic situations. *Journal of Cognitive Engineering and Decision Making, 1*(1), 22-55.

O'Connor, P., Flin, R., & Fletcher, G. (2002). Techniques used to evaluate Crew Resource Management training: A literature review. *Human Factors and Aerospace Safety, 2*(3), 217-233.

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Science, 9*(5), 242-249. doi:10.1016/j.tics.2005.03.010

Salas, E., Wilson, K. A., Burke, C. S., & Wightman, D. C. (2006). Does crew resource management training work? An update, an extension, and some critical needs. *Human Factors, 48*(2), 392-412.

Suchy, Y. (2009). Executive functioning: Overview, assessment, and research issues for non-neuropsychologists. *Annals of Behavioral Medicine, 37*(2), 106-116. doi:10.1007/s12160-009-9097-4

ASSESSMENT OF MANUAL FLYING SKILLS BY COMBINING AIRCRAFT PARAMETERS WITH PILOT CONTROL INPUTS

Ekkehart Schubert, Björn Appel, Gerhard Hüttig
Institute of Aeronautics and Astronautics
Section Flight Guidance and Air Transportation
Technische Universität Berlin
Berlin, Germany

In a recent full flight simulator experiment, 57 commercial airline pilots (long-haul captains and short-haul first officers) had to accomplish a realistic scenario ending in a manual flown "raw data approach" (localizer and glideslope capturing with absence of autopilot and flight director aid). The manual flight performance was evaluated by an instructor. Various flight path parameters and pilot control input data were collected from the simulator for an objective evaluation and comparison with the instructor assessment. At first the outer control loop was analyzed by using flight path tracking errors. This showed significant difference between the two groups of pilots. Additionally the pilot steering (roll and pitch) strategy and the aircraft reaction were analyzed in the frequency domain to analyze the inner control loop. The steering strategy for pitch consist significant lower frequency than for bank in both groups. The results showed that pilots used steering inputs with higher frequency than the aircraft reaction frequency. Pilots with lower instructor evaluation show significant higher portions in this ineffective frequency band. The results show that a combination of outer and inner loop parameters is a good indicator to evaluate pilot manual flying skills.

Introduction

Issues with manual aircraft control are a contributing factor in 42% of all accidents. 67% of these accidents would have been avoidable by training (IATA 2012). Many studies and accident reports indicate an increasing lack in manual skills, due to the increasingly automated environment; pilots are reluctant to revert to manual flying when automated systems fail. So pilots should be well trained in manual flying skills to control the aircraft especially in critical situations. During pilot training in full flight simulators pilots train difficult malfunctions (e.g. hydraulic or engine failures) and they have to demonstrate an adequate performance to meet authority requirements. The training of manual flying skills is usually done on the job, without special simulator lessons. The pilot performance is checked subjectively by an observer. To implement a training program for manual flying skills as a countermeasure into the safety management process, it is necessary to use objective and consistent performance metrics.

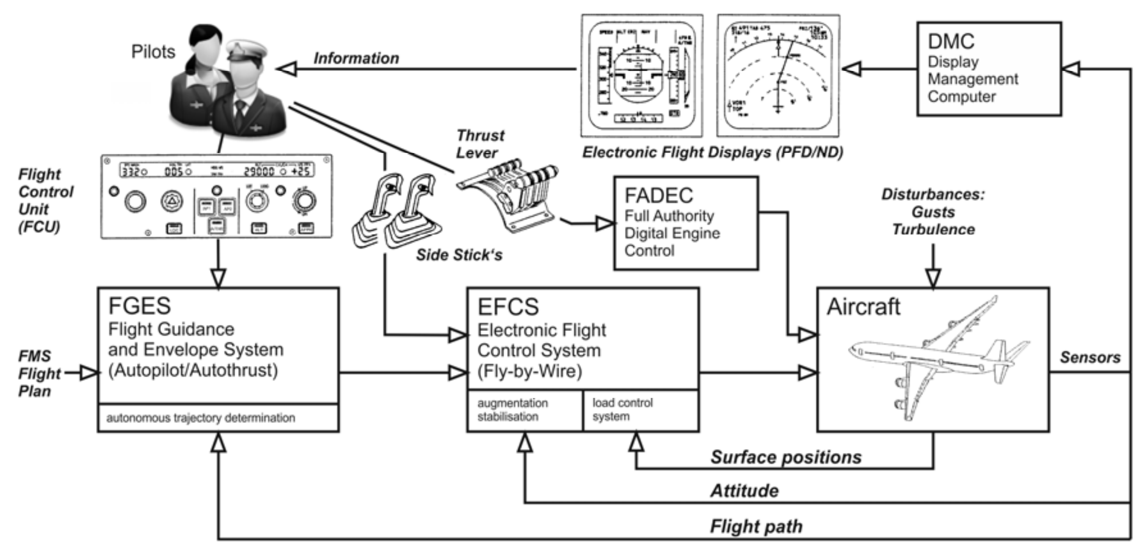


Figure 1. Illustration of Airbus A340 flight control loops (adapted from Airbus Industries 2000)

When evaluating performance of pilots on manual aircraft control, it is common to measure the outer loop control of the flight path in a time based analysis. Statistical metrics only have strong validity if applied to parameters that are associated to a well descripted flight path such as an instrument landing system (ILS). There is a certain disassociation between the control inputs of the pilot and the flight path response of modern and large transport aircraft, because of factors like inertia, transport delays of signals, possible control power and the relatively high stability of the machine. Especially in airplanes from Airbus Industries with fly-by-wire control laws, pilots have no longer direct influence on the aircraft control surfaces (see Figure 1).

From subject matter expert's opinion, the skilled pilot operation is done in a more pre-cognitive mode: To archive a smooth flight path this could be archived with less side stick inputs in a gentle manner and mostly to only one axis. Pilots with less skill level trend to give more diagonal inputs on both stick axes in a relative aggressive manner. The aim of this study is to evaluate these effects with three tiers of statistical methods.

Methods

Method 1: Time based analysis

Research studies like Mixon (1981) or Johnson (2005) describe pilot performance by analyzing various flight parameters (altitude, speed, deviation) and calculating statistic parameter out of them. In this study, the outer control loop was analyzed by using various flight parameter and flight path tracking errors. The Standard Deviation (SD) and Root Mean Square Error (RMSE) Deviation Index are good indicators to describe the preciseness. The sum of the RMSE and SD of the lateral and vertical profile is combined to the Deviation Index for both directions. The target flight path is given by the ILS.

Method 2: Frequency based analysis

The pilot steering (roll and pitch) has the main influence on the aircraft movement and trajectory. The steering strategy or inner-loop can be analyzed in the frequency. The analysis of the control input strategy is less intuitive and requires more technical resources to achieve significant results. This method has the advantage to estimate the pilot performance in all phases of flight. Rantanen (2001) and Ebbatson (2009) have already tested this method successfully in previous studies. Rantanen used flight data parameter from a small aircraft and Ebbatson used a Boeing 737 Full Flight Simulator. The main results show significant differences in the used frequency bands. More skilled pilots were able to adapt their steering inputs and show smaller variance in the used steering frequency. In this study the Power Spectral Density (PSD) is calculated by the Fast Fourier Transformation (FFT) of the steering signal. The PSD is calculated in five frequency bands from 0.01 until 0.3 Hertz (Hz).

Method 3: Steering pattern based analysis

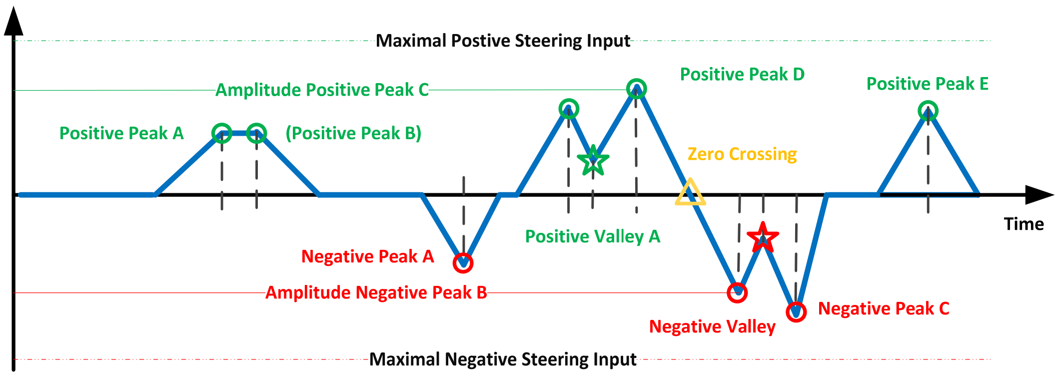


Figure 2. Steering Input Nomenclature

Under normal conditions all flight control surfaces of modern Airbus aircraft are controlled by various flight control computer and were moved by hydraulic actuators (see also Figure 1). The pilot is using a sidestick to control the aircraft in roll and pitch. The pilot commands roll rates in roll axis and g-loads in pitch axis. The control surfaces and the sidestick deflection are not proportional. Corrections in steering commands are characterized by multiple peaks and valleys in the steering signal (see Figure 2), which can be detected precisely with a signal peak detection algorithms (Appel 2012). For this study the Valley Peak Ratio (VPR) is used to describe the number of steering correction and a high VPR is an indicator of a high amount of steering corrections. The pilot can command pitch and roll simultaneously and these inputs can be measured as a Single or Dual Steering Input (SSI or DSI). The Dual-Single-Ratio (DSR) describes the proportion of number of SSI and DSI.

Experiments

This study was conducted in cooperation with a major European airline. Pilots with different levels of practice and training were selected randomly. Twenty-seven long-haul captains with Airbus A330/340 type-rating and thirty first officers with Airbus A319/320/321type-rating had participated. Two certified (JAR-STD 1A Level D) full flight simulators, with Airbus A340-600 and A320 configuration, were used for this study.

To provide a realistic scenario for manual flying tasks, a real approach was developed (for details see Haslbeck *et al.* 2012). During base leg turn, the auto flight approach mode (automatic interception of localizer and glide slope provided by the ILS) could not be armed. Shortly after this mode confusion event, the whole auto flight system was disabled and immediately the Pilot flying (PF) had to perform a raw data ILS approach. The aircraft performance, trajectory and control steering inputs were recorded with a sample rate of 15 Hz from the flight simulation process. The overall flight performance of the PF also was evaluated by an experienced instructor.

Hypothesis

Pilots with higher deviation in track and altitude use more steering inputs (H1a) and inputs with higher amplitude (H1b). Pilots with higher deviation concentrate their steering inputs on few frequency bands (H2a) and tend to higher frequencies bands (H2b). Pilots with higher deviation use simultaneously more steering inputs with on two axes (H3). Pilots with higher deviation correct their steering inputs more often (H4; Valley-Peak-Ratio). Pilots with a height deviation use steering inputs with higher frequency as the aircraft reaction frequency (H5).

Results

Method 1: Time based analysis

The deviation index showed significant difference between the two groups of pilots (A320 and A340).

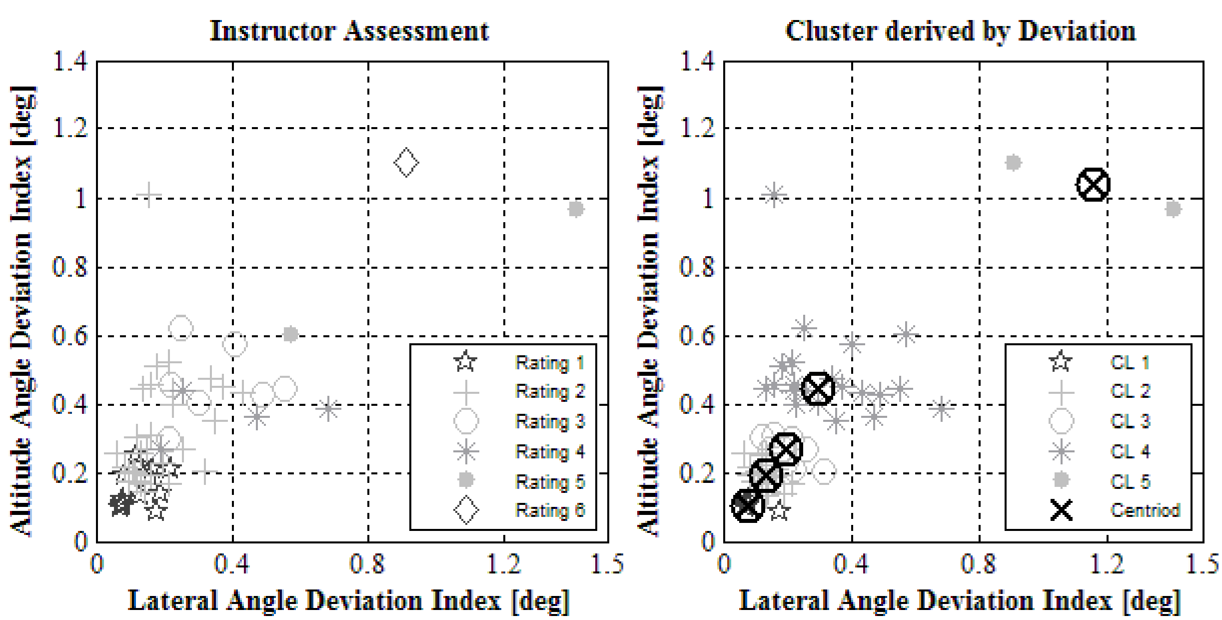


Figure 3. Instructor Assessment and Cluster derived by Deviation Index

The deviation index and the instructor ratings correlate in most cases (Figure 3, left) and a cluster algorithm is use to group similar results by using the deviation index for lateral and vertical deviation. The result of five clusters (CL) is shown (Figure 3, right). The members of CL1 represent pilots with the 5% lowest and 5% highest deviations (CL5). The clusters CL2 - CL3 represent the main group with 25-75% of the deviation index probability. The member of CL1 and CL2 are summarized to one single Cluster “good performers” and the member of cluster CL3, CL4 and CL5 are summarized to one single cluster “improvable performers” to achieve groups with approximately equal members.

Method 2: Frequency based analysis

The statistical results (see Table 1 on next page) indicates that’s the number of steering inputs (H1a) are not significantly different in both groups and only a trend to a higher number of steering inputs can be determined. The statistical results confirm the hypothesis H1b, that good performers are using steering inputs with lower amplitudes in roll. The steering strategy is analyzed by using the five deviation index cluster for pitch and roll commands. The steering command for pitch consist significant lower frequencies than in roll for all groups. The PSD results (Figure 4) indicate that pilots with lower deviation are using steering inputs in the effective frequency bands (VLF - MF). But Pilots with higher deviation concentrate their steering inputs on 1-2 frequency bands only and tend also to use steering inputs with higher frequency. The frequency based results indicates and confirm the hypothesis H2a and H2b in most frequency bands. Especially the power spectral densities for roll commands differ significantly from each other. Pilot with lower deviations use steering input in all frequency bands and the pilots with higher deviation concentrate their steering on few and higher frequency bands. The aircraft pitch and bank angle react only in the frequency bands VHF-MF (0.01-0.18 Hz) and the pilot with higher deviation show a significant higher PSD in pitch and roll in this ineffective steering area above 0.18 Hz (H5).

Method 3: Steering pattern based analysis

The results of the steering pattern analysis (Figure 4) show only tendencies and no significantly results. The results for dual input differ in roll and pitch. The pilots with lower deviation use more dual inputs in pitch to control the aircraft and neglect the hypothesis H3. But in roll pilots with lower deviation use less dual inputs and tends to confirm the hypothesis H3. The hypothesis H4 can be confirmed significantly for roll and pilots with higher deviation correct their inputs more often and show a higher valley peak ratio. In pitch the valley peak ratio is tended to be higher for pilots with higher deviations.

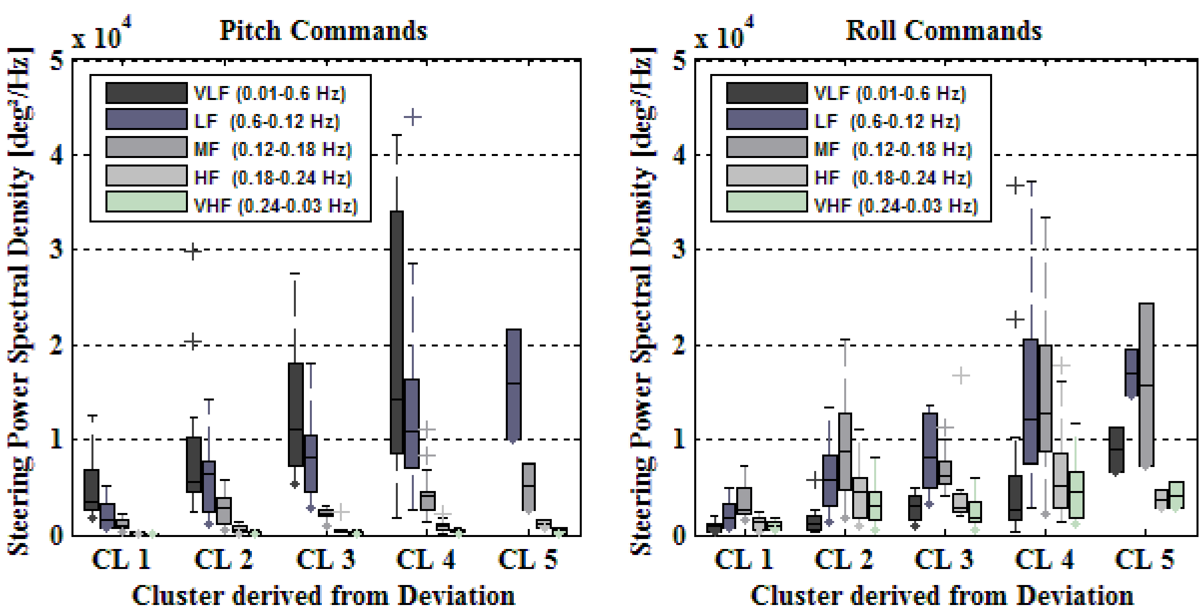


Figure 4. Steering Power Spectral Density (PSD)

Table 1.
Statistical Results.

Variable Unit	Axis	Group	No	Mean	SD	VAR	T - Test p Value	U - Test p Value	F-Test p-Value
Number of Inputs [-]	Pitch	CL good CL improvable	33 23	139,9 161,8	54,5 59,8	2971,4 3581,3	0,1622	0,1967	0,6179
Number of Inputs [-]	Roll	CL good CL improvable	33 23	184,0 197,6	54,8 53,1	3003,9 2817,6	0,3586	0,4192	0,8908
Mean Amplitude of Inputs [deg]	Pitch	CL good CL improvable	33 23	2,18 2,30	0,47 0,42	0,2 0,2	0,3316	0,4050	0,6163
Mean Amplitude of Inputs [deg]	Roll	CL good CL improvable	33 23	3,34 4,29	0,72 0,98	0,5 1,0	**0,0001**	**0,0004**	0,1069
SD Amplitude of Inputs [deg]	Pitch	CL good CL improvable	33 23	1,63 1,87	0,33 0,35	0,1 0,1	**0,0120**	**0,0137**	0,8363
SD Amplitude of Inputs [deg]	Roll	CL good CL improvable	33 23	2,27 3,03	0,61 0,67	0,4 0,5	**0,0001**	**0,0001**	0,6279
PSD VLF (0.01-0.06 Hz) [deg²/Hz]	Pitch	CL good CL improvable	33 23	2112 3492	1951 2035	3,8E+06 4,1E+06	**0,0133**	**0,0001**	0,8116
PSD LF (0.06-0.12 Hz) [deg²/Hz]	Pitch	CL good CL improvable	33 23	1253 1274	980 574	9,6E+05 3,3E+05	0,9272	0,2790	**0,0110**
PSD MF (0.12-0.18 Hz) [deg²/Hz]	Pitch	CL good CL improvable	33 23	949 1359	622 849	3,9E+05 7,2E+05	**0,0420**	0,0555	0,1057
PSD HF (0.18-0.24 Hz) [deg²/Hz] No ACFT Reaction	Pitch	CL good CL improvable	33 23	723 1056	802 582	6,4E+05 3,4E+05	0,0947	**0,0038**	0,1196
PSD VHF (0.24-0.3 Hz) [deg²/Hz] No ACFT Reaction	Pitch	CL good CL improvable	33 23	674 1391	631 1125	4,0E+05 1,3E+06	**0,0036**	**0,0002**	0,0029
PSD VLF (0.01-0.06 Hz) [deg²/Hz]	Roll	CL good CL improvable	33 23	1775 6414	1416 8287	2,0E+06 6,9E+07	**0,0026**	**0,0007**	**0,0000**
PSD LF (0.06-0.12 Hz) [deg²/Hz]	Roll	CL good CL improvable	33 23	6252 17642	4093 15815	1,7E+07 2,5E+08	**0,0002**	**0,0001**	**0,0000**
PSD MF (0.12-0.18 Hz) [deg²/Hz]	Roll	CL good CL improvable	33 23	7438 14145	4568 7844	2,1E+07 6,2E+07	**0,0002**	**0,0015**	**0,0053**
PSD HF (0.18-0.24 Hz) [deg²/Hz] No ACFT Reaction	Roll	CL good CL improvable	33 23	3859 6119	3325 4393	1,1E+07 1,9E+07	**0,0327**	**0,0215**	0,1477
PSD VHF (0.24-0.3 Hz) [deg²/Hz] No ACFT Reaction	Roll	CL good CL improvable	33 23	2678 4608	1978 2808	3,9E+06 7,9E+06	**0,0038**	**0,0063**	0,0691
Dual/Single Steering Input Ratio [-]	Pitch	CL good CL improvable	33 23	1,3604 1,1372	2,1896 1,4244	4,7945 2,0290	0,6695	0,6891	**0,0388**
Dual/Single Steering Input Ratio [-]	Roll	CL good CL improvable	33 23	1,0697 1,1682	1,2494 1,4146	1,5610 2,0011	0,7844	0,6526	0,5115
Valley/Peak Steering Input Ratio [-]	Pitch	CL good CL improvable	33 23	0,2920 0,3267	0,0862 0,0940	0,0074 0,0088	0,1590	0,0909	0,6433
Valley/Peak Steering Input Ratio [-]	Roll	CL good CL improvable	33 23	0,2870 0,3529	0,0813 0,1052	0,0066 0,0111	**0,0107**	**0,0439**	0,1809

Discussion

The results of this study show, that frequency and steering pattern methods can be used to evaluate pilot performance independently from aircraft parameters and they are suitable on fly-by-wire controlled aircrafts. The PSD analysis present similar results as in other studies and the metric can be further developed to achieve an objective pilot measurement metrics. The steering pattern method shows heterogeneous results by evaluating steering performance and can be improved to give pilots a better feedback on their manual flying performance. This metric is easier to understand and is more helpful to implement corrective action during pilot training sessions.

Acknowledgements

This work was funded by the German Federal Ministry of Economics and Technology via the Project Management Agency for Aeronautics Research within the Federal Aeronautical Research Program (LuFo IV). The authors thank all participants of the experiment for their contribution to the project.

References

Airbus Industries (2000). *A330 Flight Crew Operating Manual (FCOM)*. FCOM/OEB for LTU – Lufttransport Unternehmen. Toulouse (France): Airbus Training & Flight Operations Support Division.

Appel, B.; Schubert, E., Hüttig, G. (2012). *Assessment of Manual Flying Skills in Full Flight Simulators*. Retrieved online from EATS 2012 proceedings http://halldale.com/files/halldale/attachments/Appel.pdf.

Ebbatson, M. (2009). *The Loss of Manual Flying Skills in Pilots of Highly Automated Airliners* (PhD Thesis). UK Cranfield: Cranfield University.

Haslbeck, A., Schubert, E., Onnasch, L., Hüttig, G., Bubb, H., & Bengler, K. (2012). Manual flying skills under the influence of performance shaping factors. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 41(Supplement 1/2012), 178–183.

IATA (2012). *Man and Machine - Understanding the human factor will be essential when training the next generation of aviation professionals*. Retrieved online from http://www.iata.org/publications/airlines-international/april-2012/Pages/training.aspx.

Johnson, N., Rantanen, E. (2005). Objective Pilot Performance Measurement: A Literature Review and Taxonomy of metrics. *13th International Symposium on Aviation Psychology*.

McDowell, E. D. (1978). The Development and Evaluation of Objective Frequency Domain Based Pilot Performance Measures in ASUPT.

Mixon, T. R.; Moroney, W. F. (1982). *An Annotated Bibliography of Objective Pilot Performance* Measures. Orlando.

Rantanen, E., Johnson, N., (2001). *The Effectiveness of a Personal Computer Aviation Training Device, a Flight Training Device, and an Airplane in Conducting Instrument Proficiency Check*, Volume 2: Objective Pilot Performance Measures.

REDUCING AGGRESSIVE RESPONSES TO TCAS: EVALUATION OF A TCAS TRAINING PROGRAM

Elizabeth S. Fleming, Amy R. Pritchett, Vlad M. Popescu, William P. Cleveland
Georgia Institute of Technology
Atlanta, GA

The Traffic alert and Collision Avoidance System (TCAS) is an aircraft collision avoidance system designed to prevent mid-air collisions. While responding to a TCAS advisory is generally the safe course of action, instances of overly aggressive responses have resulted in injuries to crew members and passengers as well as disruptions in air traffic operations. However, current training standards do not address the need to mitigate overly aggressive responses. This paper details the design and evaluation of a training program for TCAS which incorporated a learning objective related to mitigating aggressive responses to advisories. The impact of the training program was evaluated by comparing the results of two flight simulator experiments. These experiments examined "trained" and "untrained" pilot responses to TCAS advisories in an integrated flight deck-Air Traffic Control simulator. Overall, the training program had a significant impact on the pilots' behavior and aggressive responses to TCAS advisories were decreased.

The Traffic alert and Collision Avoidance System (TCAS) is an aircraft collision avoidance system designed to prevent mid-air collisions. TCAS delivers a two-stage advisory: the first stage, the 'Traffic Advisory' (TA), provides an initial alert to direct the pilot's attention, while the second, the 'Resolution Advisory' (RA), displays a vertical advised collision avoidance maneuver. During an advisory, danger is imminent, and TCAS is assumed to have better, more up-to-date information than the ground operated air traffic control (ATC) facility. Thus, following a TCAS RA is generally the safe course of action.

To prepare pilots to respond to TCAS advisories, training standards for TCAS, as outlined in Federal Aviation Administration (FAA) Advisory Circular 120-55C, explicitly separate the training requirements into two segments: ground-based training requirements (e.g. classroom or computer training) and flight training requirements (e.g. simulator-based) (FAA, 2011a). This type of division is common in aviation training and aligns with training for other systems. The areas that must be covered in ground training are primarily related to general concepts of TCAS and its operation. These standards highlight the need for pilots to understand the types of advisories TCAS delivers as well as how those advisories are generated. While, TCAS flight training standards require that pilots be provided the opportunity to maneuver in response to a TCAS advisory at some point in their simulator training. Typically, TCAS flight training is integrated with line-oriented flight training (LOFT) and may be encountered within a sequence of other events.

One important concern not included in current (ground or flight) training requirements is overly aggressive pilot responses. Maneuvers in response to corrective TCAS RAs should be initiated with an acceleration (i.e. pull-up or push-down) of 0.25 g to then track the commanded vertical speed (FAA, 2011b). Responding to a corrective TCAS RA should typically cause an altitude deviation of no more than 300 to 500 feet with vertical speeds that are not excessive (FAA, 2011a). However, instances of overly aggressive responses to TCAS RA's have resulted in injuries to crew members and passengers as well as disruptions in air traffic operations. One example of this is the response of Far Eastern Transport Flight EF306 in 2009 to a Descend RA. In this instance, the aircraft was maneuvered into a dive which at one point exceeded 12,000 feet per minute, resulting in the injury of twenty crew and passengers (Aviation Safety Council, 2007). Subsequently, a bulletin released by Eurocontrol in 2011 discussed the need for training objectives intended to prevent excessive responses (Eurocontrol, 2011).

Similarly, a human-in-the-loop study conducted by the Cognitive Engineering Center at Georgia Tech found pilots averaging an altitude deviation of approximately 750 feet for Climb RAs and approximately 1,200 feet for Descend and Crossing Descend RAs (Pritchett et al, 2012). Likewise, the TCAS Operational Performance Assessment (TOPA) program at MIT Lincoln Labs found similar pilot responses to TCAS advisories. TOPA monitored the occurrence of TCAS RA's in the terminal area of eight major airports and found a maximum altitude deviation of 1,400 feet (Olszta et al, 2011). Altitude deviations of this magnitude could potentially disrupt air traffic operations, as another flight path may be located 1000 feet above or below the pilot's cleared path.

This paper discusses the development and evaluation of a training program intended to train pilots to understand TCAS use for collision avoidance in the actual traffic and operational environment. Specifically, this paper evaluates the impact of this training on pilots' aggressiveness in their responses to TCAS RAs.

Training Program Design

For initial knowledge training followed by task engagement, the complement of two training techniques was used for this TCAS training program: *Demonstration Based Training* (DBT) and *Event Based Training* (EBT). Additionally, this structure allowed an organization similar to that called for in FAA training guidelines with a division between ground-based (e.g. classroom or computer training) and flight training (e.g. simulator training). An added time constraint limited the combined program length to less than fifty minutes, reflecting the duration of time available in actual training programs to cover any one aircraft system.

DBT is a method of training through which the learner acquires knowledge, skills, and attitudes through various instructional features, including passive guidance and support as well as the observation of demonstrations of task performance (Rosen, et al., 2010). For the TCAS training program, DBT was implemented using a computer-based program designed in Microsoft PowerPoint and lasted approximately twenty-five minutes.

Also lasting approximately twenty-five minutes, the EBT segment was conducted using an aircraft simulator and presented traffic events to the pilot that created the requirement to act within a realistic environment. Feedback on pilot performance in each training event was facilitated by a researcher posing as the flight instructor and was based on pre-determined performance measures. If the pilot did not successfully meet any particular performance measure, the flight instructor reviewed the correct response for that scenario.

The traffic events were implemented within a two-crew flight deck based on the Boeing 747-400, emulated by the 'Reconfigurable Flight Simulator' (RFS) software (Ippolito & Pritchett, 2000). TCAS was emulated according to the standards required of the Minimum Operational Performance Standards (MOPS) for TCAS, including providing a TSD (RTCA, 1997). In each training and experimental scenario, the pilot participant acted as Captain and was assisted by a researcher posing as the First Officer (FO). The FO performed the duties of the 'Pilot Monitoring,' including communicating with ATC and reviewing standard checklists. Another researcher posed as the air traffic controller and provided commands to the participant via simulated radio for which the pilot wore a standard flight deck headset. The air traffic controller also communicated with other aircraft in the simulated airspace representing party-line information, created by a third researcher that the participant was able to overhear.

The training objectives for this TCAS training program follow FAA mandated training standards and also address common TCAS misunderstandings. Specifically, these objectives include general TCAS knowledge, TCAS advisory logic, the mental rules that should be invoked when responding to TCAS advisories, and the signs, signals, and symbols that the pilot can reference about the traffic to properly interact with TCAS. A training objective related to reducing aggressive response features was incorporated in both the DBT and EBT segments, however it is not included in FAA training standards. In DBT, a directed lesson was given about the need for compliance without excessive maneuvering as well as the potential consequences of excessive RA responses. In EBT, instructor feedback on pilot performance was related to both compliance and aggressiveness, including the need to reduce the aircraft's vertical rate with a weakening RA.

Simulator Study

The effectiveness of the TCAS training program was carefully evaluated through the duration of the DBT and EBT training phases as well as at the completion of the training program through pilot responses to six experimental (or 'data') flights. A prior study conducted in January 2012 serves as a baseline for comparison with the trained pilots in this study. Both the baseline and the trained pilots had been previously trained for TCAS by their carrier; the term 'trained pilots' is used here to denote pilots who completed the modified training program.

Sixteen pilots participated in the baseline study. All of the participants were male, ranging in age from their mid-20's to 59 years old. Eight held the rank of Captain in their airline, seven were ranked as First Officers, and one did not respond to the question (Pritchett et al., 2012). Eighteen pilots participated in the training study, recruited using a nearly identical recruiting protocol to the baseline study. All of the participants were male, ranging in age from their mid-20's to 59 years old. Four of the participants held the rank of Captain and fourteen were ranked as First Officers.

In both the baseline and training studies, the pilot's task was to fly a Standard Arrival Route (STAR) acting as the Captain. Typically, the flights began around an altitude of 10,000 to 20,000 feet and lasted fifteen minutes. The flights ended during the approach intercept, i.e. when the aircraft was within 'one dot' of the localizer beam indicating the approach course. The weather was calm with no wind. However, Instrument Meteorological Conditions (IMC) applied for the duration of the flight, as there were no out-the-window visuals; thus, the pilot could only reference a traffic situation display and air traffic communications for information about the traffic situation and could not visually acquire a target.

The pilots who participated in the training study flew six 'data' flights, with two traffic events in each flight. Some of events resulted in only TCAS TAs and required no maneuvering. The more severe events resulted in TCAS RAs. The run order of the scenarios was varied between pilots using a Latin Square design to compensate for possible run order effects. Three fixed variables defined each traffic event: the information provided to the pilot about traffic prior to the TCAS event (i.e., ATC Information), traffic density, and the TCAS advisory created by the target aircraft's trajectory. The ATC information about the traffic was varied at three levels: Callout, Party-line, and Conflicting. For events with a traffic callout, the air traffic controller gave a call-out to the pilot about the other aircraft prior to the TCAS event. In events containing party-line information, the background radio chatter from other traffic contained relevant information about the target aircraft but may or may not have been recognized by the pilot. In those events with conflicting guidance, the air traffic controller instructed the pilots to "Descend for traffic" moments before the pilot received a TCAS advisory to Climb. The traffic density was varied at two levels: light and heavy. The density represented a subjective measure of how congested the airspace appeared and was simulated by ATC communications with other traffic occurring at a much higher rate in heavy traffic density, creating the appearance that the controller was extremely busy.

Of interest in this paper, the aggressiveness of pilot responses to TCAS RA's can be evaluated using four measures of effectiveness: Altitude Deviation Over Duration of RA, the Average Vertical Rate Difference, the Maximum Vertical Rate Difference, and the Maximum Vertical Rate. After completing the training program, measures of aggressive RA responses are expected to decrease. Figure 1 depicts a graphical example of how the vertical rate measures can be viewed; the measure Altitude Deviation Over Duration of RA is the integral of the pilot's vertical rate.

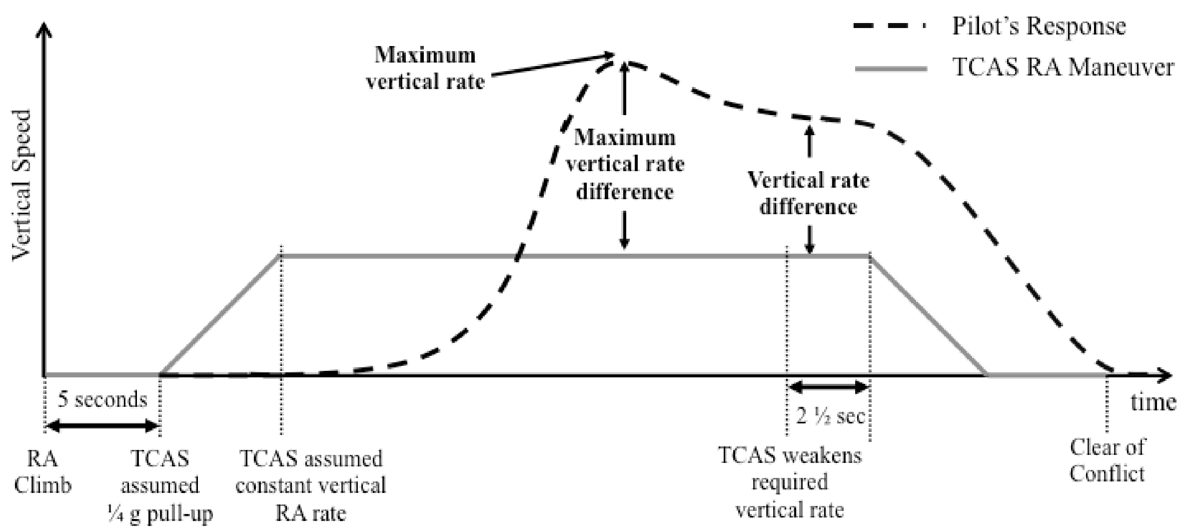


FIGURE 1. TCAS maneuver, pilot response, and corresponding measures of aggressiveness

Several other measures were also collected, including measures of compliance, recorded interaction with ATC, and pilot opinions gathered through questionnaires. However, their analysis is outside this paper's focus on aggressive responses and will be documented elsewhere.

Results

Pre-Training Results

Before completing the training program, each pilot was asked to complete a brief 'Pre-Training TCAS Quiz' comprised of eleven multiple choice questions related to material taught by the training program. Pilot responses to the Pre-Training Quiz concluded that only two of the eighteen pilot participants (11%) correctly knew that an RA should typically cause less than 500 feet of altitude deviation.

During Training Results

In the training events, the pilots had significantly lower Altitude Deviation Over Duration of RA and the Average Vertical Rate Difference in two of the training events (Climb and Crossing RA's) compared to the baseline study, as shown in Figure 2. The pilots were also more consistent during the EBT in the Climb RA event in the measure Altitude Deviation Over Duration of RA and in the Conflicting Guidance event in the measure of the Average Vertical Rate Difference.

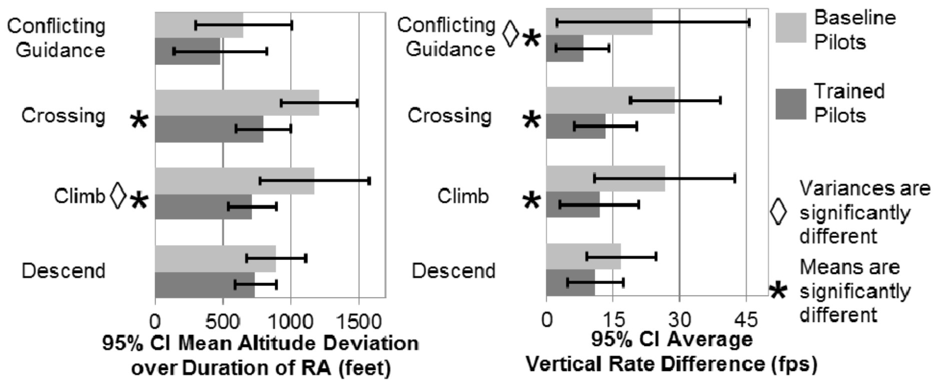


Figure 2. Mean and 95% confidence interval of the measure Altitude Deviation Over Duration of RA and Average Vertical Rate Difference within each EBT event, comparing pilot responses during training to prior baseline study

Post Training Results

Examining the measures taken in the experiment after the training, aggressive responses were reduced as indicated by nearly all measures for most events, when compared to the baseline study, as shown in Figure 3 and Figure 4. In many cases, the responses were also more consistent within the trained pilots compared to the baseline, i.e. the variance was also significantly lower.

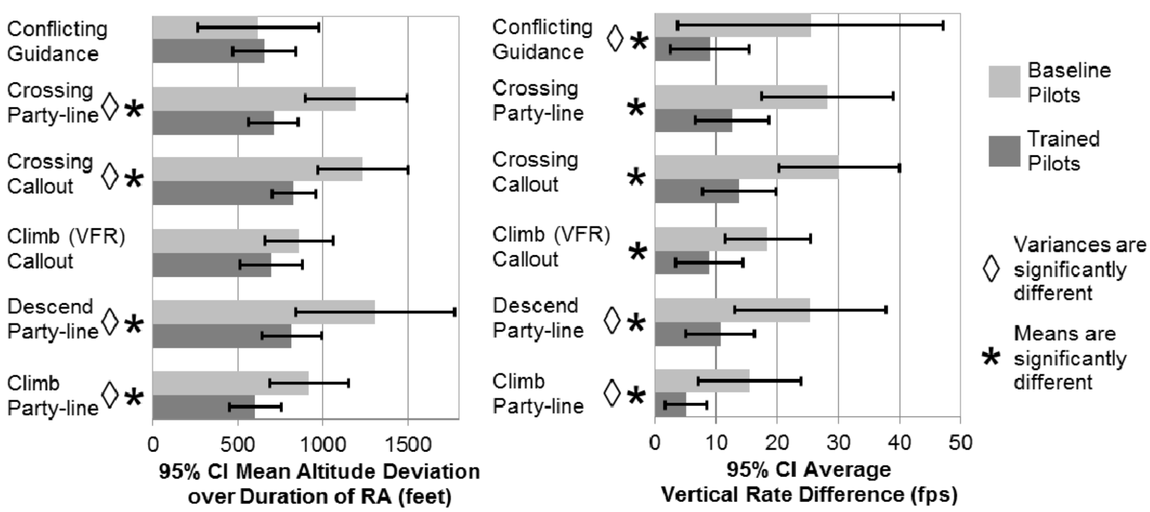


Figure 3. Mean and 95% confidence interval of the measure Altitude Deviation Over Duration of RA and Average Vertical Rate Difference within each experiment event, comparing trained pilot responses to prior baseline study

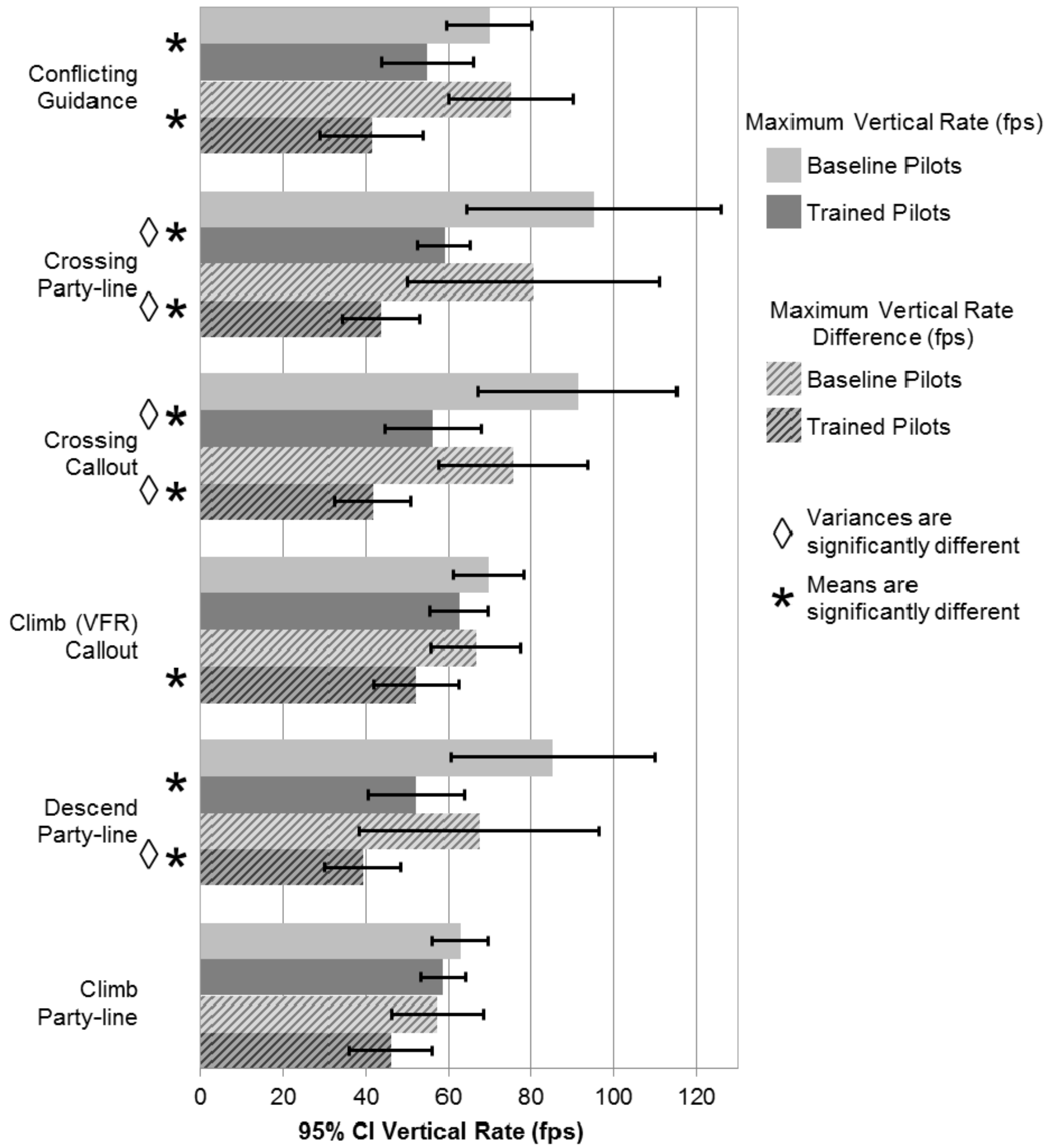


Figure 4. Mean and 95% confidence interval of the Maximum Vertical Rate Difference and Maximum Vertical Rate within each experiment event, comparing trained pilot responses to prior baseline study

Conclusions

This research adds to the discussion regarding current TCAS training objectives outlined by the FAA. As evident through reported injuries caused by eager pilots, response to TCAS RAs should be sufficient but not overly aggressive. However, current training objectives do not include the need for preventing aggressive responses. Results of the study presented in this paper concluded that aggressive responses to TCAS RAs could be significantly decreased through the inclusion of a training objective related mitigating aggressive responses to RAs.

Current TCAS training objectives reflected in FAA Aircraft Circular (AC) 120-55C, Appendix 6 include language related to understanding the various RA types and thresholds. However, language is also needed to stress the potential negative impacts of aggressive responses. Additionally, language could be added about the criteria for issuing a weakening RA; i.e. weakening RAs are given when a safe vertical separation is achieved before a safe range separation and are intended to minimize the aircraft's altitude deviation. For the training objectives related to flight training, language is already included focusing on pilot responses to weakening RAs. However, in the implementation of TCAS flight training, instructors should ensure the need to reduce the aircraft's vertical speed with a weakening RA is stressed appropriately.

Acknowledgements

The authors are grateful for the time of the pilots who assisted, specifically the thirty-four pilots who acted as participants. Additionally, the authors would like to acknowledge Henry Tran, Jack Ridderhof, Alyssa Whitlock, and Justin Mullins for their assistance in running the experiment.

This material is based upon work supported by a Cooperative Agreement (DTFAWA-10-C-00084) with the Federal Aviation Administration (FAA) Human Factors Research and Engineering Group, with Tom McCloy acting as technical manager. The authors also gratefully acknowledge the time and expertise provided by Wes Olson of MIT Lincoln Labs as a technical consultant.

References

Aviation Safety Council. (2007). Far Eastern Air Transport Flight EF306, Boeing 757-200/Thai International Airways Flight TG659, Boeing 777-300 Near Collision at an Altitude of 34,000 Ft. and 99 NM South of Jeju Island, Korea on November 16, 2006. Taipei, Taiwan.

Eurocontrol. (2011). Focus on Pilot Training (Vol. 12). ACAS II Bulletin.

Federal Aviation Administration (2011a). Advisory Circular No. 120-55C: Air Carrier Operational Approval and Use of TCAS II. Washington, D.C. Retrieved from http://www.faa.gov/documentLibrary/media/Advisory_Circular/AC%20120-55C.pdf.

Federal Aviation Administration (2011b). Introduction to TCAS II Version 7.1. Washington, D.C. Retrieved from http://www.skybrary.aero/bookshelf/books/1927.pdf.

Ippolito, C.A. & Pritchett, A.R. 2000. Software architecture for a reconfigurable flight simulator, *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, Denver, CO.

Olszta, J., Olson, W., McNamara, D., & Javits, L. (2011). Pilot Response to Traffic Alert and Collision Avoidance Resolution Advisories in the U.S. National Airspace System: Lincoln Laboratory, Massachusetts Institute of Technology.

Pritchett, A., Fleming, E., Cleveland, W., Popescu, V., Thakkar, D., & Zoetrum, J. (2012). *Pilot's Information Use During TCAS Events, and Relationship to Compliance to TCAS Resolution Advisories*. Paper presented at the 56th Annual Meeting of the Human Factors and Ergonomics, Boston, MA.

Rosen, M., Salas, E., Pavlas, D., Jensen, R., Fu, D., & Lampton, D. (2010). Demonstration Based Training: A Review of Instructional Features. *Human Factors*, 52(5), 596-609.

RTCA, *Minimum Operational Performance Standards for Traffic Alert and Collision Avoidance System II (TCAS II) Airborne Equipment*. 1997.

FLIGHT DECK MODELS OF WORKLOAD AND MULTI-TASKING: AN OVERVIEW OF VALIDATION

Christopher D. Wickens & Angelia Sebok
Alion Science and Technology: Boulder, Colorado

We review 24 computational modeling efforts of pilot multi-task performance and workload, to describe the manner in which they model three different aspects of pilot performance: the complexity of effort, the complexity of time management and the complexity of multiple resource interference. We then discuss the degree of validation of these models, and the validity of the context in which they are validated.

The next generation of the airspace program in the US (NextGen) calls for a variety of new concepts of operations as well as technologies, such as self separation, data-linked messages, closely spaced parallel operations and so forth (FAA, 2012). Naturally it is critical that such features preserve safety, even as they will increase efficiency. Traditionally, assessments of safety and efficiency of new technology and procedures is accomplished by human-in-the-loop simulation, yielding results based on response time, errors, workload, and more recently, situation awareness (Strybel et al., 2013). However it is also well understood that such HITL simulations can be extremely time consuming, and often lacking in statistical power because of a small sample size.

A complementary approach we advocate in this paper is the use of pilot or controller **computational models** of human performance that can be used to estimate the level of performance that will be achieved in certain conditions (Foyle & Hooey, 2008; Pew & Mavor, 1999; Gray, 2007). To the extent that *such models are valid* (an issue we address extensively in this paper), they can provide satisfactory predictions of performance in a fraction of the time, and at a fraction of the cost of full HITLs.

This paper describes one part of a project we performed for the FAA to establish the state of the art of computational models of pilot (e.g., flight deck) performance. What studies have been done? What aspects of performance have been modeled? And how valid are these models? In all, we identified 187 separate **modeling efforts**: typically a separate research paper that describes the application of a model to pilot performance. Many of these articles described the same **model architecture**, as applied to two or more different applications, or sets of PITL data. We then classified these efforts in terms of 13 different **aspects** of pilot performance, such as situation awareness, pilot error, pilot-automation interaction, and so forth. Details of each of these can be found in our full report (Wickens Sebok et al, 2013). However the current paper focuses only on two closely related aspects of pilot performance models: workload and multitasking. The reason for this restriction is twofold. (1). Workload and multitasking issues are of critical importance in the flight deck, as satellite navigation and improved sensors are transferring more responsibilities and tasks from ground to the fight deck. (2) These areas capture a long standing theoretical and practical interest of the first author, in their applications to aviation (e.g., Wickens Goh et al., 2003. Wickens, Sandry & Vidulich, 1983; Wickens & McCarley, 2008).

Model validation. It is our position that the best measure of model validation reflects the ability of a model to accurately predict performance (including measures like workload or situation awareness) across a **set** of flight conditions (e.g., NextGen [Ng] vs conventional [C] procedures, with advanced [A] versus older [O] equipment), so that the **differences** or variance between such conditions is accurately predicted. Such prediction, for the 4 condition case described above, is represented in figure 1, and is best captured by the product-moment correlation (r) between model predictions and P(Pilot)ITL data. While r describes the overall success of the model, the graphic scatter plot (e.g., figure 1) provides additional information concerning which conditions may be over- or under-predicted by the model. Furthermore, while the correlation coefficient may be the benchmark or gold standard of validation, other validation efforts short of this may still provide useful information. In the following, we do not discriminate between these different levels of validation, (but see Wickens, Sebok et al., 2013).

Workload and Multi-tasking. The concepts of workload and multi-tasking are closely related, both relating to the limits of the pilot's information processing capacity, but also distinct (Wickens & McCarley, 2008). Mental workload, generically relates to the relation between the total **demands** on that capacity imposed by single, or by multiple tasks, and the **availability** of cognitive resources to meet that demand. While higher workload may often diminish performance, it does not necessarily do so, if demands do not exceed operator capacity; hence effective measures of workload are often found, not in performance measures, but in physiological or subjective measures (Wickens & Yeh, 1988).

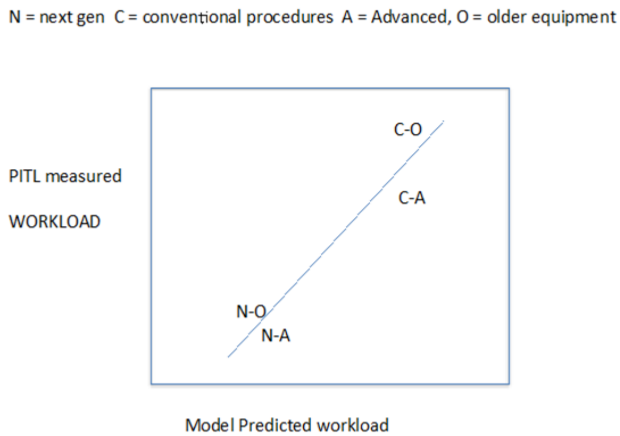


Figure 1: Correlational scatter plot representation of model validation of four predicted conditions. The regression line would suggest a high correlation (r >0.90) and hence strong validity.

In contrast, while performance will depend partially on the relationship between resources available and demanded (but only **when the latter exceeds the former),** there are other factors that may degrade performance, particularly when the latter depends on different (multiple) resources within human information processing. For example a pilot who must simultaneously activate two controls with the same hand may not experience high workload; but dual task performance of the tasks depending on the two controls will suffer (see Wickens Sebok et al, 2013).

Results: Workload and multi-tasking model validation.

Altogether 24 modeling efforts were identified that focused on either the workload or multitasking aspect (some of these may have also focused on other aspects as well), and most of these contained empirical PITL data in their write up (although these data were often not adequate to be considered true validation). We represent these efforts in the context of Table 1 which represents the multi-dimensional array of model types. Table 1 contains three columns. Each column is arrayed on a continuum of increasing model complexity or sophistication from bottom to top. The dimension of complexity are separate for single resource models of effort (left column) and of time (center column) and for multiple resource models (right column). However here we note that greater complexity typically requires greater sophistication, specialization and training of the model user; in a way that often inhibits the ubiquity of the model, even as it might also avail greater precision of model predictions. We also note that a given model may populate more than one column, being for example, simple on one column, but more complex on others.

Within each column, each number represents a unique modeling effort, whose identity can be found in part A of the reference list. Each effort (number) is in turn associated with two additional attributes. We have made an effort to classify the extent to which the modeling effort is **validated** or not, characterized by the underline of the ID number, and the extent to which we consider the modeling effort carried out on tasks or within a context that may be considered **high fidelity or validity to transport aircraft** (e.g., cockpit automation, use of line pilots in validation data) or not, characterized by the **bold face** coding for higher fidelity. Thus a modeling effort coded by both attributes (e.g., **67**) would be considered particularly valuable to the FAA.

Table 1: Three aspects of model complexity increasing from bottom to top. Each model effort is coded by a number, whose full citation is found in part A of the reference list. **Bold faced**: high fidelity. Underlined: validated.

		Demand-resouce interacion
		Weighted additive conflict be resources
		5 **14** 15 23 24
Computational model of cognitive complexity	Formal models: (ACT-R)	"VACP" (Overload/channel) (
1 11 **18**	1 5 12 16 17	**4** 15
		"VACP" summed over chann
	Simple queuing models	**4** 15
Effort (SWAT/TLX/Bedford, etc.)	10 22	"VACP" Table Lookup
Pilot assigned or SME assigned	Task overlap?	**2** **3** **4** 19 **20**
13	7 15	
	Time (TR/TA) TLAP	**MULTIPLE RESOURCES**
	11 21	
SINGLE	RESOURCES	

At the base/foundation of the table, we consider the simplest view of attention or processing capacity as a limited single resource. At this fundamental level, there are two different conceptions as to what this resource may be. On the left, it is considered to be a limited "pool" of processing effort (Kahneman, 1973), that has a physiological basis in brain metabolism (Parasuraman & Rizzo, 2007; Wickens, Hollands et al, 2013, chpt 11), and can be assessed by physiological measures such as heart rate variability, or by subjective rating scales. In the center column, the single limited resource may be considered **time**, and hence workload may be characterized by the ratio of time required to time available; and performance breakdowns related exclusively to the extent to which the latter former exceeds the latter. We now describe the increasing complexity(from bottom to top) within these columns in more detail.

Single Resources: Effort. Single resource models, where resources are based on the concept of effort may, on the one hand, simply have as inputs, pilot or SME assigned values of the effort of certain tasks (e.g., a SWAT or TLX rating). However these become somewhat circular to the extent that model validation is itself based upon a subjective workload assessment measure. More valuable, but more complex, are models in which predicted workload is computed on the basis of some objective computational algorithm of cognitive, perceptual or motor complexity (e.g., Boag et al, 2007) directly related to mental workload. For example in [18], a model of FMS complexity is derived, based on number of elements (modes) and interrelationship between them.

Single Resources: Time. When time is considered the single resource for which all tasks compete, there are many more modeling efforts that have varied in their level of complexity of the treatment of predicted time allocation between tasks that are not performed concurrently (see center column, table 1). At the simplest level, Parks & Boucek (11) have developed an aviation time-line-analysis procedure (TLAP) that simply tallies the total time demand of all tasks within an interval and divides this total by interval length to compute a predicted workload level. Such a technique can be made a bit more complex (and accurate) to the extent that amplified penalties are assigned proportional to the time that two tasks must make demands for the same (**overlapping**) period of time, such as when a pilot must lower a landing gear while attending to an ATC communication.

A layer of complexity is then added to the extent that models depend on formal queuing theory for their scheduling of tasks [22], and penalties are assigned for tasks that must wait until they are "served" by the single server queue (the pilot). Finally, at what we consider the highest level of complexity are those models that depend upon a formal well-validated (often in non-aviation domains) architecture, such as ACT-

R, or GOMS, or a derivative thereof. Such models are often applied in aviation to predict other aspects than workload (See Wickens Sebok et al, 2013); but they typically do include some time-management routine that will dictate the sequence in which tasks are performed, the period in which a task may be "neglected:, and hence an implicit measure of multiple task performance breakdown.

Multiple resources. While single resource models may be extended either in the direction of effort or time, the third dimension of complexity is to consider that resources are not single at all. Depicted in the right column, instead the pilot's information processing system contains separate resources such as visual versus auditory perception or vocal vs manual responses. To the extent that two tasks rely upon separate resources, they will be more successfully time shared (although will not necessarily lower workload). Thus the pilot will be more likely to hear an auditory warning signal if she is simultaneously reading a data link message (visual) than hearing an auditory ATC communication. The identity of these multiple resources is typically based on a fundamental model developed by Wickens (1984) that defines resources along three (now four) dichotomous dimensions (See Wickens Hollands et al, 2012 for a current version). Because these separate resources were originally assigned to four categories by model developers (visual, auditory, cognitive and psychomotor; Aldrich et al, 1989), we generically refer to this as the "VACP approach", even as the actual complexity of what defines resources is sometimes increased and may vary slightly between applications (see text below).

Fundamental to all such models is that tasks are assigned by a SME, model runner or "table lookup" to one or more resource types (e.g., comprehending ATC instructions will be considered an "A-C" task within the VACP model). Then within each resource a **demand level** is selected, often on a 1-7 scale. Some models stop at this point; others create a simple workload prediction by summing these values across channels, and at a still greater level of complexity some examine the extent to which any single channel is "overloaded" (e.g., cognitive value >5) by the sum of all task demands within that resource. Such overload predicts a multiple task performance breakdown. Finally, most faithful to the original multiple resource models are those that assign **weighted penalties** to task pairs to the extent that they share more dimensions **within the multiple resource space.** (Wickens, 2008). Thus for example two linguistic perceptual tasks will create more conflict (higher predicted penalty) than a linguistic and a spatial perceptual task.

Discussion: Validity, validation and complexity.

Tallying the two codes assigned in Table 1, we reach the conclusion that 15 of the 24 model efforts have been validated; but of those 15, only four appear to be carried out in what might be described as a high fidelity context. Correspondingly, of the 14 high fidelity studies, only 4 report careful validation.; and the reader should be aware that we have defined these two criteria liberally. If we were to restrict "high fidelity" to true NextGen concepts, and "validation" to use of correlations the number would be reduced substantially.

We also note the paucity of high fidelity validation toward the top of the table, where complex models predominate. Indeed two of these four only address cognitive complexity of component aspects of the flight deck. We account for this state of affairs in more detail in Wickens, Sebok et al, (2013), but note here the difficulty of accomplishing full validations of complex models, with the limited resources often made available for modeling efforts. However in this regard, we note and emphasize that *every modeling effort need not be validated*. Once a model **architecture** is validated in one context, greater faith can be held, that its un-validated predictions will nevertheless be accurate in a different context, hence allowing the great shortening of the time required to assess the viability of NG technology and procedures that we discussed at the outset of this paper.

References

A. Numbered references in table 1

[1] Gil, G.H., & Kaber, D. (2012, in press) An Accessible Cognitive Modeling Tool for Evaluation of Pilot–Automation Interaction. *International Journal of Aviation Psychology, 22.*

[2] Gore, B.F. & Corker, K.M., (2000a). Human Performance Modeling: Identification of Critical Variables for National Airspace Safety. *Human Factors and Ergonomics Society Annual Meeting Proceedings*. Santa Monica, CA: HFES.

[3] Gore, B.F. & Corker, K.M., (2000b). Value of Human Performance Cognitive Predictions: *Human Factors and Ergonomics Society Annual Meeting Proceedings*. Santa Monica, CA: HFES.

[4] Gore, B. F., Hooey, B. L., Socash, C., et al.. (2011). *Evaluating NextGen closely spaced parallel operations concepts with human performance models* (HCSL, Trans.) HCSL Technical Report (HCSL-11-01). Moffett Field, CA: NASA Ames Research Center.

[5] Gore, B.F. (2013 in prep). *The MIDAS User's Manual*. Moffett Field, CA: NASA Ames Research.

[7] Laudemann, I & Palmer, E. (1995). Quantitative analysis of observed workload in the measurement of aircrew performance. *International Journal of Aviation Psychology 5,* 187-197.

[8] Lyall, E.A. & Cooper, B., (1992). The Impact of Trends in Complexity in the Cockpit on Flying Skills and Aircraft Operation. *Human Factors Society Annual Meeting Proceedings*. Santa Monica, CA.

[9] Manton, J.G., Hughes, P.K. (1990). Aircrew tasks and cognitive complexity. Paper presented at the *First Aviation Psychology Conference*, Scheveningen, Netherlands..

[10] Muraoka, K. and H. Tsuda (2006). Flight Crew Task Reconstruction for Flight Data Analysis Program. *Proceedings of the Human Factors Society Annual Meeting 50*(11): 1194-1198.

[11] Parks D. & Boucek ,G. workload prediction: challenges. In McMillan, G.R., Beevis, D et al.. *Applications of human performance models to system design* . New York City, NY: Plenum Press

[12] Polson, P.G., & D. Javaux (2001). A model-based analysis of why pilots do not always look at the FMA. *Proceedings of the 11th International Symposium on Aviation Psychology*. Ohio State.

[13] Rickard, W. W., & Levison, W. H. (1981). Further Tests of a Model-Based Scheme for Predicting Pilot Opinion Ratings for Large Commercial Transports. *Proceedings of the 17th Annual Conference on Manual Control*, pp. 247-256.

[14] Riley, V., Lyall, E., Cooper, B., & Wiener, E. (1991) *Analytic methods for flight-deck automation design and evaluation. Phase 1 report: flight crew workload prediction*. FAA Contract DTFA01-91-C-00039. Minneapolis Minn: Honeywell Technical Center.

[15] Sarno, K. & Wickens, C. (1995). The role of multiple resources in predicting time-sharing efficiency:. *International Journal of Aviation Psychology, 5*(1), 107-130.

[16] Schoelles. M.J., Gray, W.D. (2011). Cognitive Modeling as a Tool for Improving Runway Safety. *T Proceedings of the 16th International Symposium on Aviation Psychology*. Dayton, OH. 541-546.

[17] Schoppek, W. & Boehm-Davis, D.A. (2004). Opportunities and Challenges of Modeling User Behavior in Complex Real World Tasks. *MMI-Interaktiv*, 7, June. 47-60. ISSN 1439-7854.

[18] Sebok, A., Wickens, C., ,et al.. (2012). The Automation Design Advisor Tool (ADAT):. *Human Factors and Ergonomics in Manufacturing and Service Industries. 22*(5), 378-394.

[19] See, J.E. & Vidulich, M.A. (1998). Computer Modeling of Operator Mental Workload and Situational Awareness in Simulated Air-to-Ground Combat:. *The International Journal of Aviation Psychology*. 8(4), 351-375.

[21] Stone, G., Culick, R. & Gabriel, R. (1987) Use of task timeline analysis to assess crew workload. In A. Roscoe (Ed) *The practical assessment of pilot workload*. NATO AGARDograph #282.

[22] Walden, R.S., Rouse, W.B. (1978). A Queueing Model of Pilot Decision making in a Multitask Flight Management Situation. *IEEE Transactions on Systems, Man and Cybernetics*. Pp.867-875, December

[23] Wickens, C.D., Harwood, K., et al.B. (1988). TASKILLAN: A simulation to predict the validity of multiple resource models of aviation workload. *Proceedings of the 32nd Meeting of the Human Factors Society.* Santa Monica, CA: Human Factors Society. Pp.168-172.

[24] Wickens, C.D., Larish, I. & Contoror, A. (1989). Predictive Performance Models and Multiple Task Performance. *Proceedings of the Human Factors Society 33rd Annual Meeting*, pp 96-100.

B. References cited in text.

Aldrich, T.B.,et al.. (1989) The development and application of models to predict operator workload during system design. In G. MacMillan et al (Eds.), *Applications of human performance models to system design* (pp. 65080). New York: Plenum.

Boag, C., Neal, A., Loft, S., & Halford, G.S. (2006). An analysis of the relational complexity in an air traffic control conflict detection task. *Ergonomics, 49*, 14, pp 1508-1526.

FAA (2012, March). *NextGen Implementation plan*. Washington D.C: Federal Aviation Administration.

Foyle, D.C., & Hooey, B.L. (2008). *Human Performance Modeling in Aviation*. Boca Raton, FL: CRC Press

Kahneman, D. (1973) *Attention & effort*. Englewood Cliffs N.J.: Prentice Hall.

Strybel, T, et al (2013) Measuring the impact of NextGEN Operating concepts on Situation Awareness & Workload *Int J. Aviation Psych. 23,* 1-26.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63-101). New York: Academic Press

Wickens, C. D. (2008b). Multiple resources and mental workload. *Human Factors Golden Anniversary Special Issue*, 3, 449–455.

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W., & Talleur, D. A. (2003). Attentional models of multi-task pilot performance using advanced display technology. *Human Factors, 45*(3), 360-380.

Wickens, C. Hollands, J. Banbury S.& Paraasuraman R.(2012) aEngineering Psychology & Human Performqnxe 4th edition. Upper Saddle River N.J.: Pearson

Wickens, C.D. & McCarley, J.S. (2008). *Applied Attention Theory*. New York: CRC Press, Taylor & Francis Group.

Wickens, C. D., Sandry, D., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, output, and central processing. *Human Factors, 25*, 227-

Wickens Sebok, A. Keller, J. et al (et al. (2013). Modeling and Evaluating Pilot Performance in NextGen-. FAA final report DTFAWA-10-X-800. Annex 1.11

Yeh, Y. & Wickens, C.D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors* 30(1): 111-120.

Acknowledgments

This research was sponsored by grant #DTFAWA-10_X-800 from the Federal Aviation Administration to AlionScience, via NASA Ames Research Center. The authors gratefully acknowledge Dr David Foyle of NASA and Dr. Brian Gore of San Jose Research Institute for their scientific/technical oversight of this work; and of John Keller for his contributions to the research team. The opinions expressed in this report are solely those of the authors.

TRACKING AND VISUOSPATIAL WORKING MEMORY

Robin Langerak, BSc.
Visualization and Simulation (VSIM) Centre, Carleton University
Ottawa, Canada
Chris M. Herdman, Ph.D.
Scientific Director VSIM Centre, Carleton University
Ottawa, Canada
James Howell, BA.
VSIM Centre, Carleton University
Ottawa, Canada

The present research examines the role of visuospatial working memory in supporting pursuit tracking. Participants completed a pursuit motor tracking task while simultaneously completing secondary tasks designed to separately place demands on either storage or processing in visuospatial working memory. The results show that simple pursuit tracking utilizes visuospatial processing in working memory without a strong requirement for visuospatial storage. These findings have implications for understanding time-sharing of tasks in the cockpit.

The ability to track moving objects is critical to supporting a number of tasks in the cockpit, including but not limited to, maintaining a flight path, pursuing other aircraft, as well as monitoring and maintaining flight parameters such as an aircraft's attitude. Research has demonstrated that object tracking requires substantial cognitive resources (Pylshyn & Storm, 1988) and a link has been made between tracking and visuospatial working memory (Baddeley & Logie, 1999). It has also been shown that performance on pursuit motor tracking tasks is sensitive to interference from other cognitive tasks (Mastoianni & Schopper, 1986). The objective of the present research was to examine what aspects of visuospatial working memory are involved in pursuit motor tracking.

Working memory is used to temporarily store and manipulate information and has been described in terms of a multicomponent model that includes a central executive, a phonological loop, and a visuospatial sketchpad (Baddeley & Hitch, 1974; Baddeley & Logie, 1999; Baddeley, 2011). The central executive is a limited capacity system that controls the allocation of the attentional resources that support task performance. The phonological loop and visuospatial sketchpad are limited capacity storage and processing systems for verbal and visual information, respectively. A large body of research has examined the phonological loop. This research has shown that the phonological loop can be divided into two distinct subsystems: a phonological store and an articulatory rehearsal (processing) mechanism (Baddeley, 2001). It has been suggested that tracking is supported at the visuospatial sketchpad component of working memory (Baddeley & Logie, 1999) and that the sketchpad can also be subdivided into two subsystems: a visual cache for storage and an inner scribe for processing visuospatial information (Logie, 1995).

In pursuit motor tracking, participants are required to locate and monitor the changing position of a moving target while generating and executing a series of motor commands to maintain alignment with the moving target. This task requires constant online processing of the visual target and is therefore assumed to involve the visuospatial sketchpad's inner scribe subsystem. The inner scribe is thought to process information about spatial relationships and motion (Logie, 1995). Since tracking does not require participants to store information about a moving target's previous locations, pursuit tracking is viewed as a relatively pure processing task that places little, if any, demand on the visual cache (storage) subsystem.

In the present research, participants performed a pursuit motor tracking task in baseline and dual-task conditions. Two dual-task conditions were used. In one of the dual-task conditions tracking was combined with a visuospatial *storage* task. In the other dual-task condition, tracking was combined with a visuospatial *processing* task. It was hypothesized that relative to baseline performance, tracking performance would decrement more when combined with the visuospatial processing task than with the visuospatial storage task.

Method

Participants. Twenty-three undergraduate university students participated in the experiment. Participants had normal or corrected-to-normal vision. Three participants were dropped from the final analysis because their performance was found to be at or below chance for one or more of the experimental blocks, resulting in a final sample of 20 participants.

Materials. The target was a 5x5 moving grid (300x300 px) in which a single moving dot (processing task) or a pattern of dots (storage task) was presented. During pursuit tracking conditions a participant-controlled cursor consisting of a red-hashed outline of the grid was displayed (See Figure 1). Stimuli moved according to sinusoidal functions for X and Y motion such that the grid moved in a pseudorandom figure 8 pattern. The target grid moved at a maximum speed of 300 px /sec on the x- and y-axes while the participant-controlled cursor moved at a maximum of approximately 1360 px/sec on the x-axis and 650 px/sec on the y-axis. The cursor was purposefully programmed to move faster than the target to ensure that participants could catch up with the target.

Cursor inputs and storage/processing task responses were collected with an Xbox 360 controller connected to a Dell XPS computer. The experiment was coded in C++ using OpenGL and SDL libraries and was presented on an Asus V678H 120Hz 27-inch LCD monitor. Tracking performance was recorded as the Euclidian distance between the target grid and the cursor, sampled at 10 Hz. Performance on the visuospatial tasks was recorded as button-press responses on the direction pad of the game controller. Joystick sensitivity was set at a fine grain to ensure that the tracking task would be challenging - even for individuals with considerable experience using the controller.

Tasks. The experiment consisted of three baseline (single-task) conditions (tracking, visuospatial storage, visuospatial processing) and two dual-task conditions (tracking + visuospatial storage, tracking + visuospatial processing). Each condition consisted of forty 15-second trials, each of which were initiated at participants' press of the right joystick button. All blocks were preceded by practice trials.

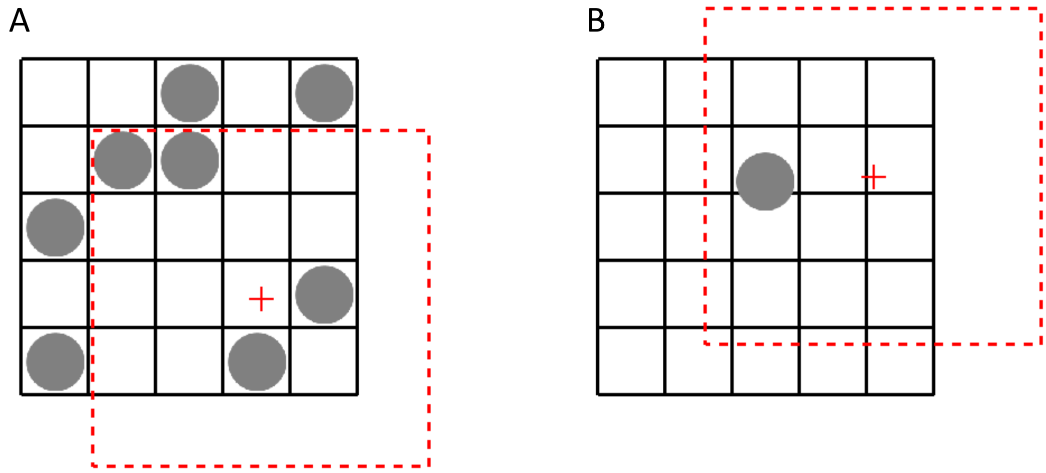


Figure 1. Depictions of the 5x5 target grid and the red participant-controlled cursor used in the dual-task conditions: Storage (A) and Processing (B). In the single-task tracking condition, no dots were displayed in the target grid (not depicted).

#

For the tracking task, participants were instructed to use the controller's right joystick to position the cursor such that it overlapped the target grid. Each trial started with both the target and the cursor aligned at the center of the screen. After a 500ms delay the target would begin moving in one of four possible directions and participants would be attempt to follow it with the cursor. For the first 1000ms the target speed was increased exponentially from 0 to its maximum speed to allow participants to accommodate.

In the single-task visuospatial storage condition eight dots were presented for 1500ms in randomly selected cells of the 5 x 5 target grid and were then removed. Participants were instructed to maintain a visual image, or 'mental snapshot', of the dot pattern during a 7-second retention interval. Following the retention interval a second eight-dot pattern was presented for 1500ms and participants were to indicate via a button-press response whether this was the same or different from the first pattern. On half of the trials the second dot pattern was identical to the first pattern. On the other half, one dot had been moved by one cell. The dual-task visuospatial storage condition was identical to the single-task condition described above, except that participants had to track the target while simultaneously performing the storage task.

The processing task was specifically designed to index visuospatial processing without placing any demand on storage. In the single-task visuospatial condition, participants were required to monitor the up-down motion of a single dot located in the target and to make button-press responses each time the dot changed direction. The up and down keys of the controller direction pad were used. Participants were to quickly and accurately make up-down responses that followed the directional changes of the dot. At the start of each trial the dot was stationary for 1500-2000ms, after which it would begin moving either upwards or downwards. The dot would continue in one direction of motion for 1000-2000ms moving at a rate of 100 px/sec and changed direction 7-10 times per trial. To prevent participants from anticipating a directional change, the dot would either switch direction somewhere within the borders of the grid, or it could roll through the upper or lower border of the grid and wrap around to the opposite side to continue its direction of motion. The dual-task visuospatial processing condition was identical to the single-task condition described above, except that participants had to track the target while simultaneously performing the processing task

Procedure. Participants completed an informed consent form and were given detailed instructions and a practice block before each of the five experimental conditions. Participants repeated the practice trials prior to each condition until they attained at least 60% accuracy. Each participant began with the baseline tracking task, followed by one of the baseline (single-task) visuospatial tasks and then the corresponding dual-task condition. After completing the baseline and dual-task conditions for one of the visuospatial tasks, participants would complete the other baseline visuospatial task and its corresponding dual-task condition. The order of the visuospatial task conditions was counterbalanced such that half of the participants received the storage condition before the processing condition, with the other half receiving the reverse order.

Results

Tracking performance was measured as Root Mean Square Error (RMSE) of the Euclidean distance between the moving target and the participant-controlled cursor. The lines of data from the start of each trial to the point where the target attained maximum speed were discarded. A one-way ANOVA with 3 levels (Condition: baseline tracking, dual-task tracking + visuospatial storage, and dual-task tracking + visuospatial processing) was used to analyze the tracking data. There was an overall effect of Condition, $F(2,38) = 12.92$, $MSE = 108.01$, $p < .001$, $\eta_p^2 = .41$, (see Figure 2). Paired samples *t*-tests showed that tracking error was greater in the dual-task visuospatial processing condition than in both the baseline tracking, $t(19) = 4.23$, $p < .001$, and the dual-task visuospatial storage conditions, $t(19) = -3.70$, $p = .002$. Tracking error did not differ between baseline tracking and dual-task visuospatial storage conditions, $t(19) = 1.02$, $p = .320$.

A full analysis of the visuospatial task performance is not included here. However, performance on the visuospatial tasks did not compromise the interpretation of the tracking results. In brief, performance on the visuospatial storage task did not differ between the baseline (single-task) and dual-task conditions. In contrast, performance on the visuospatial processing task was worse in the dual-task condition than in the baseline (single-task) condition.

Discussion

Tracking has long been used as an in-lab analog to investigate aircraft flight performance, where both tasks require constant attention and information updates about object motion in conjunction with online motor planning and correction. The present research examined the nature of the demands that pursuit motor tracking places on visuospatial working memory. Using the multicomponent model of Working Memory as a framework (Baddeley, 2001; Baddeley & Hitch, 1977), a tracking task was combined with working memory tasks designed to place

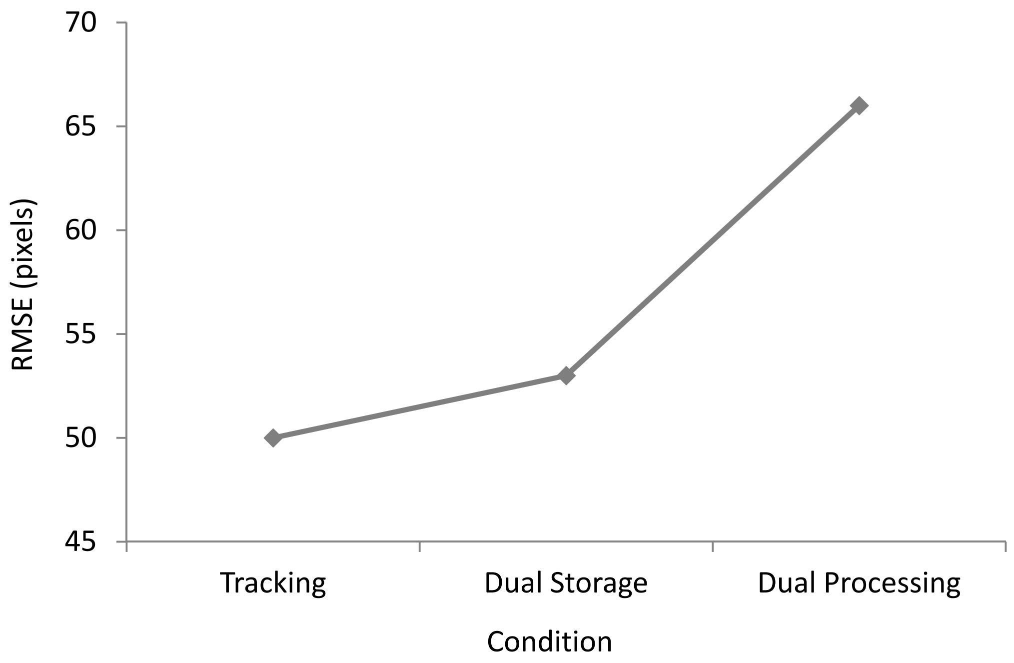


Figure 2. RMSE tracking performance (in pixels) between the target and the cursor as a function of Condition (baseline tracking, dual-task tracking + visuospatial storage, and dual-task tracking + visuospatial processing).

specific demands on either visuospatial storage (remembering a dot pattern) or visuospatial processing (detecting the directional change of a dot).

Participants were able to simultaneously perform the tracking task in combination with the visuospatial storage task without incurring any performance decrements relative to baseline performance levels. In contrast, tracking performance decremented significantly from baseline when combined with the visuospatial processing task. Because pursuit tracking is a continuous processing task, it can be considered as a visuospatial analog to articulatory suppression. Studies using articulatory suppression to specifically load the articulatory loop have provided evidence for a distinct and separable storage and processing subsystems in phonological working memory (see Baddeley & Hitch, 1994). By using pursuit tracking to continuously load the inner scribe, the present findings represent some of the strongest behvaioural evidence for Logie's (1995) suggestion of a delineation of storage and processing in the visuospatial sketchpad.

Following the assertion of distinct storage and processing subsystems, the present findings suggest that certain in-flight task that utilize the visuospatial processing component of the visuospatial sketchpad may interfere with a pilot's performance on tasks requiring pursuit tracking in the cockpit. Tasks that compete for resources in the inner scribe of the sketchpad will interfere with a pilot's ability to maintaining a flight path or flight formation, pursue other aircraft, as well as monitor and maintain flight parameters. In contrasts, tasks that utilize the visuospatial storage component of the visuospatial sketchpad are not likely to interfere with a pilot's performance on pursuit motor tracking in the cockpit.

References

#
Baddeley, A.D. (2001). Is working memory still working? *European Psychologist,* 7, 85-97.

Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In G.A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press

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.

Logie, R.H. (1995). *Visuo-spatial working memory*. Hove, UK: Erlbaum.

Mastroianni, G.R. & Schopper, A.W. (1986). Degradation of force-loaded pursuit tracking performance in a dual-task paradigm, *Ergonomics*, 29:5, 639-647

Pylyshyn, Z.W. & Storm, R.W. (1988) Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vis*ion 3, 179–197
#
#

ENHANCING MILITARY HELICOPTER PILOT ASSISTANT SYSTEM THROUGH RESOURCE ADAPTIVE DIALOGUE MANAGEMENT

Felix Maiwald and Axel Schulte
Universität der Bundeswehr München (UBM),
Institute of Flight Systems (LRT-13), 85577 Neubiberg, Germany
{felix.maiwald, axel.schulte}@unibw.de

Assistant systems investigated today are beneficial in principle, but may induce additional load for the pilot, especially if the system intervenes at a time when the human has no more free cognitive resources to adopt the offered support. This article describes an approach how to enable a knowledge-based pilots' assistant system in the domain of military helicopter missions to interact with the pilot by resource adaptive dialogue management. To minimize the automation induced additional load for the pilot, the assistant system estimates the pilots' residual mental capacity and furthermore the current cognitive workload in the first step. This assessment enables the system to direct dialogues to the perceptual modality and code, which can be assumed to provide spare resources. Within this article, we provide a description of the implemented resource model, the conducted experiments as well as results of the overall evaluation of the adaptive capabilities in a relevant mission context provided in our research flight simulator.

1. Introduction

In the domain of civilian aircraft, the rising utilization of automation has been motivated primary by technical feasibility without investigating the needs of the operator (cf. Sarter et al., 1997). As a result of this process, a variety of automated functions have been developed, which are acting more or less independently in their limited operating range without considering superior mission objectives. Monitoring, control and diagnosis of the various sub-functions is due to the human operator.

To cope with the high complexity of such automated systems and resulting handling-errors, more technical functions have been introduced in evolutionary design cycles (cf. *vicious circle*, Onken & Schulte, 2010). As a consequence of increasing complexity, the automation is neither operable nor understandable for the crew particularly in critical situations (cf. *mode confusion*, Sarter et al., 1997).

Billings (1991) specifies requirements for *"human centered automation"* thereby mitigating the appropriate described "automation induced errors". In further pursuit of Billings (1991), Onken and Schulte (2010) derived three basic requirements for such human-machine cooperation as a specification for the desired behavior of assistant systems. Schulte (2012) claims that the human operator shall perform his/her tasks by using the given operation supporting means, as long as he/she is able to do so under normal workload conditions. As long as the human succeeds without errors or excessive demands, the assistant system would not be intervening at all (region I in Figure 1). Consequently, electronic aids support the human (e.g. by interactions / interventions) in situations of discursive attention or even in periods of excessive demands (section II and boundary between region II-III in Figure 1). Here, assumptions are made that the provided support reduces the workload (WL) of the human again to a manageable level (section IV).

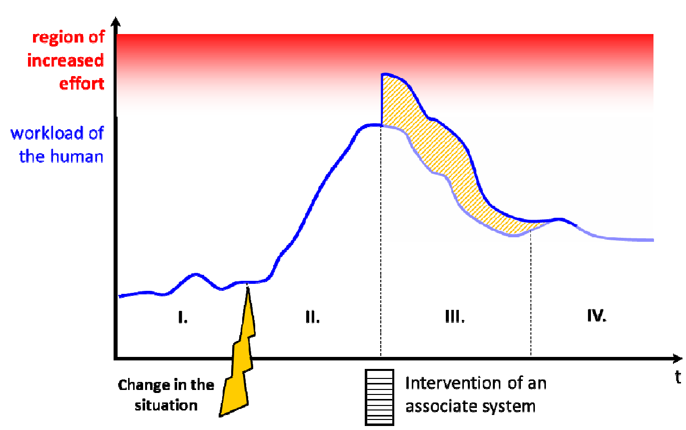


Figure 1: Desired and real relationship between interventions of an associate system and workload of the human operator

However, the human operator has to invest additional cognitive resources to recognize the offered support or interact with the system in any way. Due to the increased demand of resources, the workload level may increase at first rather than declining (boundary between region II and III in Figure 1). In extreme cases, the human is not able to provide sufficient free cognitive resources to benefit from the offered support optimally. Wiener (1989) ascribes such adverse automation induced effects to deficient human-machine-interfaces, not considering the current mental state of the operator.

To solve suchlike (automation induced) problems, an approach is presented in this article how to optimize an electronic aid in the domain of military helicopter missions. The aim of this work is to minimize the demands on additional cognitive resources, which have to be provided by the pilot to perceive and handle the offered support of an electronic aid. A vivid indicator to be minimized within this work illustrates the orange-hatched area in Figure 1.

2. MiRA, a resource adaptive associate for military helicopter pilots

In the context of future army helicopter missions, the Institute of Flight Systems of UBM developed the Military Rotorcraft Associate (MiRA) for the pilot flying. The purpose of our investigations was to enable this knowledge-based associate system to predict the human operator's workload and the remaining capacity of his/her mental resources in the current task situation. These estimations will be used to determine the optimal interaction modality in order to minimize the additional (automation-induced) demand on the resources of the pilot.

Characteristics of the Military Rotorcraft Associate

MiRA, like other known approaches of knowledge-based military pilot associate systems, e.g. CAMA (Schulte & Stütz, 1998) and the RPA (Miller & Hannen, 1999), is designed according to the principles of cognitive and cooperative automation (Onken & Schulte, 2010). This type of automation is at first characterized by the assertion that the associate system has to cooperate with the human operator in a similar way a human assistant would do. Consequently, not only the human pilot but also the associate system needs to understand the given work objective in order to derive the necessary tasks in the course of the work process. As a consequence, we developed ***task models***, incorporating the a-priori knowledge on military transport helicopter missions.

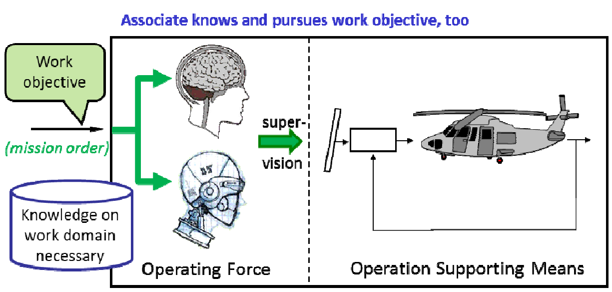


Figure 2: Work system with associate as part of the operating force

The cooperation is characterized by the basic requirements for human-automation interaction (Onken & Schulte, 2010). According to this, the associate system shall guide the operators' attention to the most urgent task if necessary. In the second chain of cooperation, the associate should manipulate the task load, to keep the subjective workload on an appropriate level. To follow this requirement, MiRA takes own initiative to start appropriate dialogs. However, dialogues as such initiated by the associate system require the pilot to provide additional mental resources. If the system does not account for the current mental resource situation, i.e. the workload of the human operator, the system-initiated dialogs might not even come through in an extreme case.

For the reasons mentioned above, we developed ***human resource models***, which enable MiRA to estimate the human operator's workload and the remaining capacity of his/her mental resources for the current task situation.

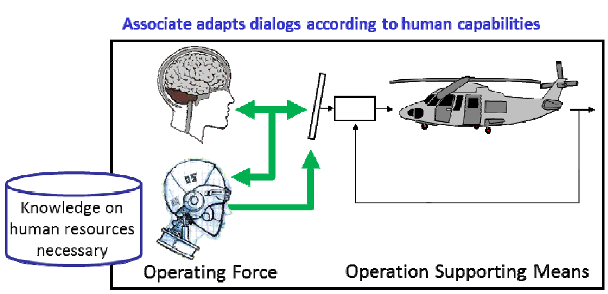


Figure 3: Associate adapts human machine interface in order to minimize additional demands on resources

This model connects information about the resource demand to each individual task. In a second step, the demands of concurrent tasks are overlaid. In general, this model assumes that the interference between several tasks is directly proportional to the predicted workload (Wickens, 2002). In a second step, the associate system proactively assesses the impact on residual operator's resources induced by additional dialogs on distinct human modalities in advance. For this purpose, MiRA adds the work demands of its intended dialog(s) estimated on the basis of the resource model, thereby going through all potential human perceptual modalities. Potential resource conflicts caused by the associate system initiating a dialogue can thereby be anticipated and, hence, prevented by selecting the modality with the lowest additional conflict, respectively mental workload.

3. Detailed concept for resource adaptive information transfer

Our implemented concept to adapt the information transfer (particularly the dialogs) according to current pilots' spare cognitive resources is summarized in Figure 4. For realization of an assistant system providing the desired capability to adapt dialogs, we build two models at first:

1.) *Model of pilot tasks* for the purpose of estimating the current tasks of the pilot
2.) *Model of pilot resource consumption* to estimate the resource consumption and WL for current tasks

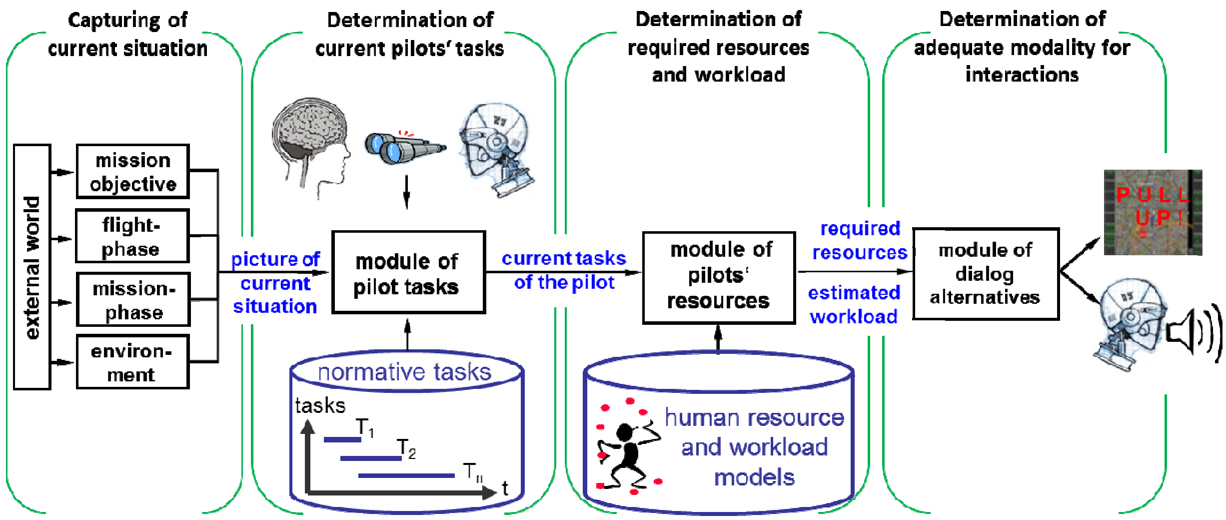


Figure 4: Concept for determination of the adequate interaction modality minimizing additional pilots' resources

Model of pilot tasks

In the first step, we capture all external influences on the pilot during a military transport helicopter mission (i.e. the state of the helicopter, the mission objective, the current flight and mission phase as well as environmental conditions). The interpretation of the helicopter task agenda leads to a pilot-specific task agenda. After aggregating this data into a full situational picture, this can be used to determine the current tasks the pilot should be executing. For this purpose, we deployed models of mission-typical task situations. These transition state networks have been developed based on knowledge acquisition experiments with German Army aviators.

To give dynamic to this normative task models, we synchronize the assumed tasks with the tasks the pilot is actually executing. Therefore, human-machine-interactions such as visual information acquisition (e.g. gazing at moving map) as well as manual interactions are analyzed. In this context, assumptions are made that observing the gazes as well as observing the manual interactions enables the assistant system to draw conclusions on the tasks actually processed by the human operator. Manual interactions that are taken into consideration include the currently displayed page on the CDU, pushed buttons, current system settings (e.g. landing gear), as well as control stick inputs. Visual interactions taken into account for this model are provided by a commercial eye-tracking system (FaceLAB®) and its integrated object-related gaze-tracking.

Model of pilot resource consumption

In a further step the determined actual task(s) are associated with our model of human resource consumption. This model is based on Wickens' (Wickens & Hollands, 2000) so called multiple-resource theory and estimates the required human resources by use of eight-dimensional demand vectors (Wickens, 2002). Every demand vector symbolizes the demand a single task poses on the human operator expressed in the terms of information acquisition, information processing and response. Hence, data were gathered through knowledge acquisition experiments, in which German Army aviators had to rate individual resource demands that arise during the different mission tasks. To eliminate subjective influences from these models as far as possible, laboratory experiments have been conducted to better match the predicted resource conflicts within distinct task situations with the objectively measured pilots' performance. Based on these experiments, we applied machine learning methods (i.e. genetic algorithms) to adapt the underlying human resource model to the measured human performance exemplary (Maiwald & Schule, 2012).

Table 1 explains the demand vectors in detail using the sample tasks "Approach H/C to Pickup-zone" and "Change zoom on map". To estimate the current resource utilization, a modified Visual-Auditory-Cognitive-Psychomotor model (VACP; Aldrich & McCracken, 1984) is used. This enables the assistant system to determine remaining available resources of the operator in case a manipulation of attention is required (first assistant system requirement from Onken & Schulte, 2010).

Table 1: *Demand vectors for 2 sample tasks*

Task	information acquisition				processing		reaction	
	visual - spatial	**visual - verbal**	**auditory - analog**	**auditory - verbal**	**cognitive – spatial**	**cognitive - verbal**	**manual**	**verbal**
Approach to Pickupzone	3	2	2	0	2	2	3	0
Change zoom on map	1	0	0	0	2	0	1	0

In addition the predictive resource consumption model inherits some metrics of task and resource conflicts for estimating current pilots' workload. For this purpose, the demand vectors of current tasks are applied to a modified workload index model (W/INDEX; Wickens, 2002) in pairs. The modified metric we applied to the W/INDEX model, eliminates any limitation on the number of tasks examinable in parallel. When considering n-tasks in parallel, we establish a quantity of k pairwise conflict values TKW_i $(i \in \{1, .., k\})$.

$$k = \frac{n!}{2*(n-2)!} \qquad (1)$$

These resulting pairwise conflict values $\mathrm{TKW_i}$ can be summarized as a row vector. $\overrightarrow{\mathrm{TKW}} = \{\mathrm{TKW_1}, \mathrm{TKW_2}, \dots, \mathrm{TKW_k}\}$. The entire estimated workload is defined according to the following formula as the geometric sum of the pairwise conflict values.

$$\mathrm{Workload} = \sqrt[2]{(\mathrm{TKW_1})^2 + \dots + (\mathrm{TKW_k})^2} = \sqrt[2]{\sum_{m=1}^{k} \mathrm{TKW_m}^2} \qquad (2)$$

Estimating desired interaction modality with lowest additional resource consumption

To minimize the additional resources, the pilot has to provide to perceive system-initiated warnings or interact with the offered support, we developed the approach depicted in Figure 5. This procedure is based on the pilots' current tasks and the derived level of workload, i.e. the utilization of resources. At first, we overlay the pilots' current tasks with hypothetical interactions (e.g. dialogs) using four different codes and modalities. Within a one-step planning process, each of these task combinations is rated by the modified W/INDEX- and VACP-resource model referring to the resulting workload level, i.e. utilization of resources. In detail, we regard the following four potential interaction channels: visual-spatial (e.g. symbolic messages on displays), visual-verbal (text messages on displays), auditory-analog (symbolic auditory messages) and auditory-verbal (speech output). The assumed utilization of each regarded resource is represented by the traffic lights (c.f. Figure 5). Finally, we derive the desired resource for an interaction in the current situation generating the lowest additional workload value (i.e. the visual-verbal resource in the example of Figure 5). In a future stage of completion, the deviation of an interaction resource will also incorporate an optimization (i.e. ensuring an equal utilization) of regarded resources.

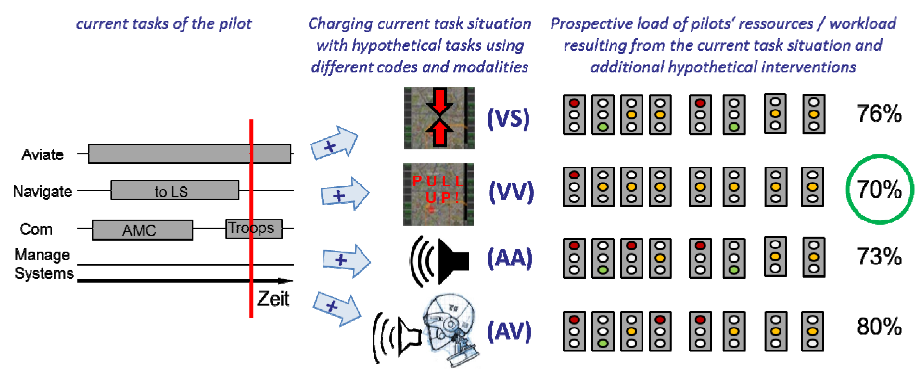


Figure 5: Estimating adequate resource for dialogs with lowest additional demand

4. Experimental evaluation of assistant system prototype

To evaluate the overall benefits of the MiRA assistant system, in particular the adaptive automation aspects, we conducted an experiment in our generic, stationary, two-seat side-by-side H/C-cockpit in 2011.

Experimental Setup

Eight German Army aviators at an average age of 37 years (min. 28, max. 51) participated as test persons. Their flying experience ranged from 830h up to 5100h with an average of 1815h. The experiments were conducted using two different experimental configurations: In the adaptive configuration the MiRA assistant system communicated with the pilot by adaptive use of either speech output, text messages, audio-alerts or symbolic display messages (8 subjects). In contrast the non-adaptive configuration was further subdivided composing dialogues either via aural text messages (4 subjects) or visual speech messages (4 subjects). Each subject participated in the adaptive as well as in the non-adaptive configuration. They were initially briefed on the nature on human-machine interfaces.

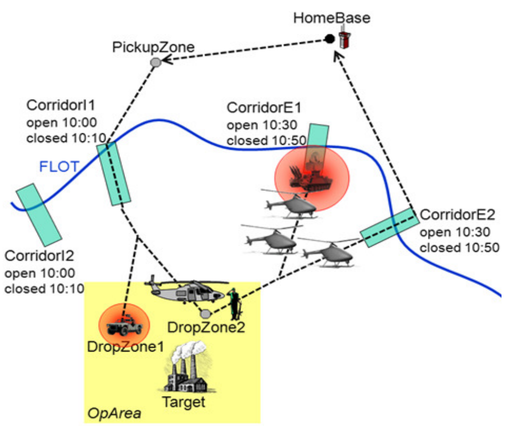


Figure 6: Reference mission for MiRA assistant system validation process

The missions lasted about 30 to 45 minutes and had the primary objective to transport troops from friendly *PickupZone* into hostile *DropZone*. It was mandatory to pass corridors with distinct opening times to transit from friendly to enemy territory and back.

The commander (represented by the pilot not flying, PNF) is entitled to control three UAVs, which support the mission by taking over reconnaissance tasks such as exploring the helicopter routes and landing sites located in hostile territory in real-time. Each mission contained a follow-up mission order that was received by the crew upon accomplishment of the primary mission (i.e. after dropping troops at Dropzone). The follow-up order contained either a second troop transport (within hostile territory) or the recovery of a crashed pilot (from hostile to friendly territory).

For the validation of the MiRA system prototype, only aspects relating to the pilot are discussed in the following chapters. Results concerning the UAV mission management by the PNF can be found in Strenzke et al. (2011). In our experiment, we required the H/C-crew to operate below an altitude of 150ft AGL in enemy territory. This altitude describes a safety-critical parameter, due to increasing exposure to enemy air defense when violating this requirement.

Results of Validation

We accumulated occurring altitude violations over the violation time as a performance parameter. This was computed for both, the *adaptive* and *the non-adaptive configuration* concerning the system generated altitude warnings. A t-test was used to compare these two configurations.

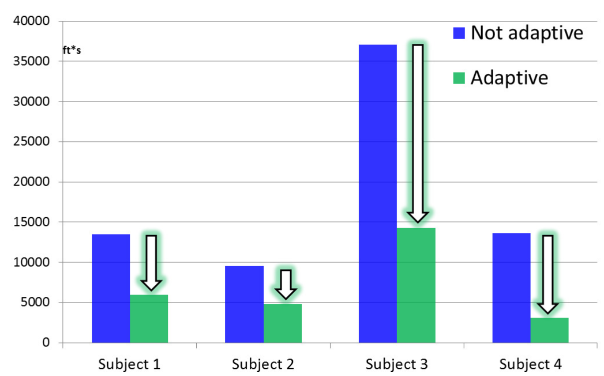


Figure 7: Comparison of altitude violation in adaptive and non-adaptive configuration.

As a result, the violation decreased for more than 50% in the adaptive configuration in comparison to the non-adaptive configuration (cf. Figure 7). That means, the performance improved significantly in our experiments (*t(4)=2.17, p=0.048, n=4* refers to the number of missions each containing approx. 20-30 minutes of low-level flight) when transferring information in a resource-adaptive way.

After each mission, questionnaires (i.e. NASA-TLX, ratings on different configurations of the associate system and on the overall system evaluation) were presented to the pilots.

To ensure comparability, all NASA-TLX-ratings were normalized due to different utilization of the workload scales. As depicted in Figure 8, pilots rated the configuration with *text messages only* as their highest subjective workload level (42.6% on the average). In contrast the *resource-adaptive-configuration* was appraised with an averaged workload level of 30.6% only. The workload reduction between these two configurations was proved significant by a two side t-test *($t(32)=2.06$, $p=0.047$, $SD=9.97$, $n_1=12$, $n_2=22$)*. In addition, the workload decreased weak significant in mean from 38.9 % in *speech-only-configuration* to 30.6% in resource adaptive mode *($t(32)=1.87$, $p=0.07$, $SD=10.2$, $n_1=12$, $n_2=22$).*

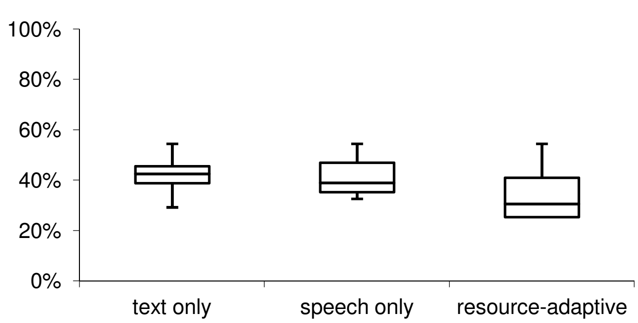


Figure 8: Pilot subjective workload norm. NASA-TLX.

Further subjective ratings regarding the specific benefit of the associate system to the pilot benefit showed a weak significant trend (t(14)=1.95, p=0.07) that pilots perceived best support in the *resource-adaptive configuration*. Comparing the configurations *resource-adaptive* vs. *text message* only, the pilots felt significantly better supported in resource-adaptive mode ($t(10)=5.06$, $p<0.001$). A comparison between *resource adaptive* and *speech-only* mode showed no significant effect.

As depicted in Figure 9, the eight pilots also rated if they experienced any difference in subjective workload between the non-adaptive and resource-adaptive system configurations. Two subjects stated no difference and six subjects attested decreased subjective workload in resource-adaptive configuration. Furthermore, the pilots attested the MiRA associate system to increase mission safety and efficiency.

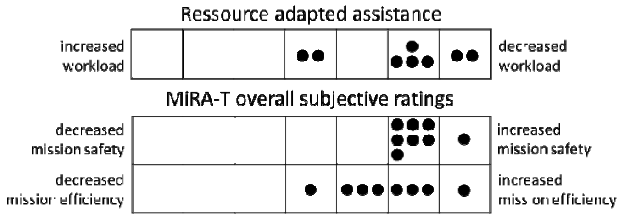


Figure 9: Overall ratings for MiRA associate system

5. Conclusions

In this article, we present an approach to enhance a knowledge-based assistant system in the domain of military helicopter missions with cooperative capabilities. For this purpose, we developed a concept for estimation of pilots' residual capacity in human resources as well as an estimation of his current cognitive workload. By the use of models considering the current resource allocation, an assistant system was enabled to transfer necessary information on remaining resources of the pilot. This proactively prevents the pilot from being overtaxed, which maximizes the performance of the overall system.

The overall MiRA associate system evaluation trials showed that the altitude-related exposure to potential threats could be significantly reduced in the resource-adaptive mode. Pilots reported decreased workload and felt best supported in the resource-adaptive information transfer configuration. Finally, the MiRA associate system was rated to be a helpful electronic crewmember increasing the mission efficiency and safety. Our future work will incorporate trials for a further validation of our resource model prototype, in particular concerning the demand vectors. Furthermore, we will apply our presented concept in the domain of civilian aircraft, i.e. emergency helicopter missions. In this context, we intend to enhance the model based task prediction by developing a hybrid approach incorporating human behavior models along the lines of the work presented in Schulte & Donath (2011).

6. References

ALDRICH, T.B. & MCCRACKEN, J.H. (1984). *A computer analysis to predict crew workload during LHX ScoutAttack Missions*. Vol.1, Fort Rucker, Alabama, US Army Research Institute Field Unit.

BILLINGS, C.E. (1991). *Human-Centered Aircraft Automation: A Concept and Guidelines*. NASA Technical Memorandum 103885. Moffet Field, NASA-Ames Research Center.

MAIWALD, F. & SCHULTE, A. (2012). Adaptation of a Human Resource Model by the use of Machine Learning Methods as Part of a Military Helicopter Pilot Associate System. *Human Factors and Ergonomics Society (HFES) 56th Annual Meeting*. Boston, Massachusetts

MILLER, C.A. & HANNEN, M.D. (1999). The Rotorcraft Pilot's Associate: design and evaluation of an intelligent user interface for cockpit information management. *Knowledge-Based Systems*, 12(8).

ONKEN, R., & SCHULTE, A (2010). *System-ergonomic Design of Cognitive Automation-Dual Mode Cognitive Design of Vehicle Guidance and Control Work Systems*. Heidelberg: Springer.

SARTER, N.B., WOODS, D.D. & BILLINGS, C.E. (1997). Automation surprises. In SALVENDY, G. (Eds.) *Handbook of Human Factors and Ergonomics* (2nd edition). S.1926-1943, New York: Wiley.

SCHULTE, A. (2012). Kognitive und kooperative Automation zur Führung unbemannter Luftfahrzeuge. *Zweiter Interdisziplinärer Workshop Kognitive Systeme: Mensch, Teams, Systeme und Automaten - Verstehen, Beschreiben und Gestalten Kognitiver (Technischer) Systeme*. Duisburg, Deutschland.

SCHULTE, A. & DONATH, D. (2011). *Measuring self-adaptive UAV operator's load shedding strategies under high workload*, 9th Conference on Engineering Psychology & Cognitive Ergonomics, HCI International.

SCHULTE, A. & STÜTZ, P. (1998). Evaluation of the Crew Assistant Military Aircraft (CAMA) in Simulator Trials. In: *NATO Research and Technology Agency, System Concepts and Integration Panel*. Symposium on Sensor Data Fusion and Integration of Human Element. Ottawa, Canada.

STRENZKE, R., UHRMANN, J., BENZLER, A., MAIWALD, F., RAUSCHERT, A. & SCHULTE, A. (2011). *Managing Cockpit Crew Excess Task Load in Military Manned-Unmanned Teaming Missions by Dual-Mode Cognitive Automation Approaches*. AIAA GNC 2011. Portland, Oregon.

WIENER, E.L. (1989). *Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft*. NASA Technical Report 177528. Moffett Field, CA: NASA Ames Research Center

WICKENS, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), pp. 159-177.

SINGLE-PILOT WORKLOAD MANAGEMENT DURING CRUISE IN ENTRY LEVEL JETS

Barbara K. Burian
NASA Ames Research Center
Moffett Field, CA

Shawn Pruchnicki
Bonny Christopher
Evan Silverman
San Jose State University
Moffett Field, CA

Carla Hackworth
Jason Rogers
Kevin Williams
Gena Drechsler
Barry Runnels
Andy Mead
FAA Civil Aerospace Medical Institute
Oklahoma City, OK

Advanced technologies and automation are important facilitators of single pilot operations, but they also contribute to the workload management challenges faced by the pilot. We examined task completion, workload management, and automation use in an entry-level jet (ELJ) flown by single pilots. Thirteen certificated Cessna Citation Mustang (CE-510S) pilots flew an instrument flight rules (IFR) experimental flight in a Cessna Citation Mustang simulator. At one point, participants had to descend to meet a crossing restriction prior to a waypoint and prepare for an instrument approach into an un-towered field while facilitating communication from a lost pilot who was flying too low for air traffic control to hear. Four participants experienced some sort of difficulty with regard to meeting the crossing restriction, and almost half (n=6) had problems associated with the instrument approach. Additional errors were also observed, including eight participants landing at the airport with an incorrect altimeter setting.

The advent of personal jets, such as entry-level jets (ELJs) and very-light jets (VLJs), has made a wider range of operations and missions available to private and professional pilots alike. Private pilots can now fly higher and faster than ever before, and commercial ventures such as air taxi operations and short charter flights are now more economical. The automation and advanced technology aboard these aircraft are essential features that make personal jet flight by single pilots possible.

However, automation and advanced technology bring their own challenges. The design of glass cockpit systems currently used in these aircraft places a heavy cognitive load on the pilot in terms of long-term, working, and prospective memory; workload and concurrent task management; and developing correct mental models as to their functioning (e.g., Burian & Dismukes, 2007). These cognitive demands have a direct relationship to pilot errors committed during flight (Dismukes, Berman, & Loukopoulos, 2007). Burian (2007) found a significant correlation between workload and time management (i.e., poor crew and single-pilot resource management, which are abbreviated CRM and SRM, respectively) and problems using advanced avionics. Additionally, almost two-thirds of the accident reports Burian analyzed involved at least one of six different cognitive performance problems (e.g., memory problems). Workload management is a crucial aspect of SRM. Best practices for single-pilot flight task and workload management must be better understood within the current operating environment, and beyond, as we move to an era of optimizing the national airspace system outlined in NextGen concepts.

In an exploratory simulation study, we examined private and professional pilot proficiency in single-pilot task and workload management using a level 5 flight training device (for simplification, a "simulator"). Participant task performance of one of the scripted high workload periods occurring during the cruise portion of an IFR flight, described below, is reviewed here. A detailed description of the entire study can be found in Burian et al. (2013).

Method

Thirteen certificated Cessna Citation Mustang (CE-510S) pilots flew an experimental flight, composed of two legs with realistic tasks in the U.S. northeast corridor, in a Cessna Citation Mustang simulator. Performance was evaluated against airline transport pilot and instrument rating practical test standard criteria (FAA, 2008, 2010), as well as the successful completion of the scripted tasks.

Participants

Six of the 13 participants fly the Mustang as professional pilots, and the other seven fly it for personal business or recreation (i.e., owner-operators). They were recruited through letters sent to the 321 airmen who possessed C510-S type ratings and lived in the contiguous 48 states of the USA at the time we conducted the study. Participants were paid a rate of $50 per hour and were reimbursed all travel costs.

Measures and Apparatus

Prior to flying the scripted flights in the simulator, participants completed three questionnaires pertaining to demographics, opinions about advanced avionics and automation, and Citation Mustang and G1000 cockpit set-up preferences. Prior to flying each leg of the experimental flight, participants were provided with standard flight bag materials and a printed flight briefing packet, including a description of the flight, proposed time of departure, aircraft location at the departure airport and planned aircraft parking at the destination airport, a departure airport diagram, a completed flight plan, a completed navigation log and weight and balance information including a weight and balance diagram, a complete weather briefing package, paper IFR en route navigation charts, complete Jeppesen Airway manuals with current paper departure, arrival, and approach plates, and airport and facilities directories.

The flight simulator, located at the FAA Civil Aerospace Medical Institute facilities in Oklahoma City, OK, featured a realistic Mustang flight deck with a G1000 avionics suite, digital control loaders, and a high fidelity digital surround sound system that accurately replicated flight, engine, system, and environmental sounds. The out-the-window display system included a 3D Perception 225 degree (lateral angle) spherical floor-to-ceiling projection screen that gave the pilot a realistic field-of-view. The G1000 default settings were adjusted to those preferred by each participant, as indicated through previously completed questionnaires. During the experimental flights, participants were asked to make instantaneous self-assessment (ISA) ratings of their workload using a small rectangular box with five numbered buttons (1 = very low workload; 5 = very high workload) when a red light at the top of the box was illuminated. Researchers controlled when the light would illuminate remotely from the experimenter's station. Once illuminated, the light stayed on for 60 seconds or until the participant pressed one of the numbered buttons. Participants were given a printed card explaining the meaning of each ISA rating for their reference in the simulator. Immediately following the completion of each leg of the experimental flight participants completed a paper-pencil version of the NASA Task Load Index (TLX) giving subjective workload ratings on each of the TLX subscales for the flight, overall, and for specific scripted high workload tasks during the leg. Following the completion of the experimental flights, audiotaped debriefing interviews were held with the participants.

Procedure

Following a review of the overall study purpose and completion of informed consent paperwork, participants were given flight briefing materials to review for a familiarization flight to be held the following day. The simulation study began the next day with introductions, calibration of the eye tracker, and the completion of a take-off and landing at KOKC. Participants then completed the familiarization flight, lasting approximately 30 minutes, from Clinton Sherman Airport (KCSM) to Will Rogers World Airport in Oklahoma City (KOKC). Participants practiced completing ISA ratings during the familiarization flight. After a brief break with beverages and snacks provided, participants were then given briefing materials for the two legs of the experimental flight and were allowed as much time as they desired to review them and prepare for the first leg. With the assistance of recently retired air traffic controller subject matter experts (ATC SMEs), who had experience managing traffic in the U.S. northeast corridor, participants completed the first leg of the scripted experimental flight (Teterboro, NJ [KTEB] to Martin State Airport in Baltimore, MD [KMTN]), lasting approximately one hour. They then completed the NASA TLX for that leg. Following a break for lunch, participants were given as much time as they desired to review pre-flight briefing materials for the second leg of the experimental flight (KMTN to Ingles-Hot Springs, VA [KHSP]), also lasting approximately one hour. At the completion of this flight, participants again completed NASA TLX measures, debriefing interviews were conducted, and participants were thanked for their participation.

Results

Demographics

In the year prior to the study, our 13 male participants reported flying the Cessna Mustang a mean of 153.7 hours (range: 68-350 hours) and flying as a single pilot in the Mustang for a mean of 138.5 hours (range: 15-350 hours). No significant differences in flight hours, experience, or self-reported skill with advanced avionics and automation were found between owner-operators and professional pilots.

High Workload Task Performance

One of the high workload periods analyzed in this study occurred approximately three-quarters of the way through the second leg of the experimental flight (KMTN to KHSP). Participants were flying at an interim cruise altitude of 16,000 ft (from an original cruise altitude at FL200) and had two major flying tasks to complete: Descend at their discretion to meet a crossing restriction of 10,000 ft. 15 nm prior to a waypoint (Montebello VOR [MOL]), and prepare for an ILS runway 25 approach into KHSP, an un-towered airport. Additionally during this period, participants could hear center controllers unsuccessfully trying to communicate with another pilot (played by one of the researchers) who was trapped under the cloud deck. Air traffic controllers asked the participants to relay communication from the "lost pilot" to them and, although participants could have declined, all 13 agreed to do so.

This high workload period lasted an average of 7 minutes and 55 seconds (SD = 31 seconds, range = 0:07:11 to 0:08:53) and ended when ATC handed the participant pilots off to another controller at the end of the lost pilot scenario, which generally occurred around the time participants crossed MOL, less than 35 nm from KHSP. Although participant completion of the approach and landing at KHSP was not part of this defined high workload period, preparation for approach and landing *was* expected to have occurred during this time. Therefore, aspects of participant approach and landing performance associated with the quality of their preparation will be discussed.

Overall Flight Performance. Although at least one error was committed by each of the participants during this high workload period or during the approach into KHSP, generally they flew the aircraft within appropriate parameters. For example, all participants maintained engine temperatures below the limit of 830^{o} and responded to all radio calls from ATC. Twelve pilots met the crossing restriction, and no excessive bank or pitch angles, or excessive yaw were observed. All participants flew close to V_{mo} (250 KIAS) during this period and their airspeeds ranged from 193 KIAS (M = 210.85 KIAS, SD = 14.51 KIAS) to 248 KIAS (M = 243.23 KIAS, SD = 4.36 KIAS) with an overall average airspeed of 228.30 KIAS (SD = 8.80 KIAS). Those flying slower airspeeds tended to be participants who reduced their speeds purposefully near the end of the lost pilot scenario apparently to increase the amount of time they had available to finish preparing for the approach at KHSP. The observable errors committed during this high workload period or the approach and/or landing at KHSP can be seen in Table 1. One professional pilot committed only one error when he neglected to report the initiation of his descent from 16,000 ft to ATC; all other participants made two or more errors during this high workload period. A one-way ANOVA revealed no significant differences between owner-operators and professional pilots with regard to the number of errors committed, $F(1,11) = 0.54$, p =.48. However, two surprising findings were the large number of pilots who neglected to contact ATC to report they had initiated their descent from 16,000 ft MSL (n = 10) and the large number who landed at KHSP with an incorrect altimeter setting (n = 8).

Crossing Restriction 15nm before MOL. Participants received the clearance to descend (at their discretion) with a crossing restriction when they were 44 nm from MOL. They began their descents when they were an average of 32.96 nm (SD = 4.44 nm) from MOL (about 18 nm from the crossing restriction point) and were traveling an average of 224.69 KIAS (SD = 10.21 KIAS). In addition to flying a descent manually, there are two primary ways to accomplish a descent using the G1000 automation: vertical speed (VS) mode and vertical path (VPTH). As its name implies, VS is an autoflight mode that is set by choosing a rate of descent in feet per minute. In contrast, VPTH is actually programmed in the G1000 and allows the pilot to indicate crossing restrictions, which the automation will then ensure the aircraft meets during the programmed descent.

Twelve of the 13 participants programmed VPTH to accomplish this task although two did not couple VPTH to the AP and just used its guidance to support their descent using VS (one of them did not make the crossing restriction—was 1,180 ft high). Additionally, due to a programming error, VPTH did not capture for one participant so he ended up using VS instead. The remaining participant used VS with no VPTH guidance as a back-up. It took the 12 participants an average of 53 seconds (*SD* = 42 seconds; range = 00:20 to 02:09) to program the VPTH descent, although there were two distinct clusters of time it took to do this programming. These clusters appeared unrelated to participant subgroup (e.g., owner-operator) or whether the VPTH was used for the descent or only for back-up information. The participants with the lowest programming times (*n* = 8, range = 20 to 38 seconds) took an average of 29 seconds (*SD* = 7 seconds) to do so; those with the longest programming times (*n* = 4, range = 01:35 to 02:09) took an average of 1 minute 58 seconds (*SD* = 10 seconds) to complete the programming. As expected, those taking more time to complete the programming interleaved other tasks while doing so.

Table 1.
Participant Errors[1] Committed.

	All Participants (n = 13)	Owner-Operators (n = 7)	Professional Pilots (n = 6)
Communication/ readback error	4	1	3
Did not report leaving 16,000 ft. MSL	10	6	4
Minor crossing restriction programming error	1	1	0
Major crossing restriction programming error	4	2	2
Failed to make crossing restriction	1	0	1
Misunderstood lost pilot comms capabilities[2]	5	4	1
Minor ILS programming error	1	0	1
Major ILS programming error	7	4	3
Landed at KHSP with incorrect altimeter setting	8	3	5
Total errors	41	21	20
Mean number of errors	3.15	3.00	3.33

[1] "Errors" includes only those that were observable
[2] Participants had difficulty understanding or remembering that the "lost pilot" could hear ATC but that ATC could not hear the lost pilot.

One participant who unsuccessfully used VS with VPTH guidance chose an initial descent rate of 2,500 fpm and continued to fly close to the maximum operating speed for the aircraft. When it became apparent that he might not make the crossing restriction, he compensated by pulling back some power but waited almost a minute before increasing his descent rate to 3,000 fpm (passing through 12,700 ft MSL, 2.58 nm from the crossing restriction point). It appeared that he knew he had not met the crossing restriction but he did not contact ATC to inform them. Two pilots, including the one who was unsuccessful, initially made an error when programming VPTH by placing the point where the crossing restriction was to be met 15 nm past MOL, instead of 15 nm before MOL. Both caught their errors fairly quickly and corrected them. To summarize, four of the 13 pilots had difficulty programming the crossing restriction or descending but only one actually failed to make the crossing restriction.

Communication Assistance for the Lost Pilot. While descending to meet the crossing restriction and preparing for their approach into KHSP, pilots assisted with transmitting communication from a lost VFR pilot to ATC. Due to problems with the simulator audio system, one participant was not presented with the lost pilot scenario during leg 2. As mentioned earlier, all the other participants agreed to assist and six volunteered before ATC could even ask. Of the pilots presented with the "lost pilot scenario," all continued to offer assistance until the situation had been resolved, with the exception of one who did not transmit the final two comms from the lost pilot to ATC because he was preparing for his approach into KHSP. Five participants had at least some initial confusion as to who could hear whom during the scenario; in those cases, the lost pilot clarified that she could hear ATC. Only one participant continued to transmit ATC comms to the lost pilot, unnecessarily, throughout the scenario.

Approach and Landing at KHSP. The lost pilot situation was typically resolved about the time that participants crossed MOL, which is 17.3 nm from the initial approach fix (IAF) for the approach into KHSP. Some participants appeared to become a bit concerned about being ready for the impending approach into KHSP during the lost pilot scenario, and five did such things as slow down or ask for vectors or some other alternate routing that would give them added time to prepare (e.g., stay on current heading a bit longer past MOL, request a different approach fix that was 5 miles further away from MOL.). Contrary to what was expected, most pilots did not actively prepare for the approach during the lost pilot scenario. Three queried ATC about aspects of the approach while assisting with the lost pilot comms (e.g., which approach could be expected), and one was observed looking at aircraft weights on the MFD, but very little of their preparation for the approach occurred during the lost pilot scenario. The other nine participants were not engaged in any observable approach preparation during the scenario.

Further analysis revealed that six pilots (three owner-operators and three professional pilots), including two who queried ATC during the scenario, had actually completed most or all of their approach preparations (e.g., reviewing/preparing for the approach) before this period or the lost pilot scenario began. Six participants could be observed entering in required frequencies into the radios quite early during the leg (e.g., on climb out from KMTN). Four participants prepared for the approach (i.e., reviewed the approach plate for the first time) between MOL and the IAF, and one professional pilot prepared very late, just before arriving at the intermediate fix. One owner-operator was never observed preparing for the approach by reviewing the approach plate prior to conducting the approach, though he did scroll down to the decision height (DH) information at the bottom of the Jeppesen chart displayed on the MFD when he was 252 ft above DH. Interestingly, of the six participants who prepared for the approach before the start of this high workload period, two actually programmed the approach at that time; the other four waited until after passing MOL, when the specific approach in use was confirmed by ATC. During the post-flight debriefings, the two who programmed the approach quite early spoke of their preference to program as soon as possible, even if it meant having to change it later. Both completed the approach without difficulty. Eleven participants programmed the approach after the end of the lost pilot scenario. The more significant difficulties encountered typically involved incorrectly programming the G1000 (e.g., not activating or arming the approach, or being in the wrong autopilot mode to capture the approach).

There was a fairly even split between those who did (n = 7) and did not (n = 6) encounter difficulty in programming or executing the approach. Not surprisingly, those who prepared for the approach quite early in the leg (n = 4, 66%) tended to have fewer difficulties programming or executing the approach than participants who completed most of their preparation activities just before conducting the approach (n = 2, 33%). Similarly, participants who programmed the approach quite early (n = 2) had no problems conducting the approach, whereas only four of the remaining 11 (36%) participants, who programmed the approach just before or even after they had begun executing it, had no problems. Interestingly, six of the nine participants who reported to ATC that they had gotten the automated weather report at KHSP, landed at KHSP with an incorrect altimeter setting (29.86 instead of 29.84), as did two others who did not check the weather prior to landing (see Table 1). The incorrect altimeter setting these eight participants landed with was the altimeter setting given to them when they descended through the transition altitude of 18,000 ft MSL much earlier in the flight, before this high workload period began.

Discussion

In this study of single pilot workload in ELJs, we found no significant differences in performance, errors made, or success rates in accomplishing the major tasks analyzed due to pilot type (owner-operator or professional pilot). It is possible that the owner-operators in our study were more experienced than most or that those with less experience or skill did not volunteer to participate. It is also possible that our professional pilots fly less frequently or are less capable than non-participants, but we have no evidence or reason to believe that this was so. The common training that all participants received through FlightSafety International, a Part 142 training center, might also account for the lack of difference found between the two groups.

Workload management when piloting technologically advanced aircraft involves the allocation of mental resources to accomplish multiple tasks concurrently. Most participants completed short tasks, such as dialing in a new altitude, before moving on to other tasks. Some participants also demonstrated a similarly focused method when programming the G1000. Almost all performed other tasks concurrently such as dialing in a new heading while listening to the rest of an ATC clearance. Most participants chose to interleave more lengthy automation programming with other cockpit tasks. Contrary to what one might expect though, those who programmed the G1000 without interruption, e.g., for the approach at KHSP or to meet the crossing restriction, made just as many programming errors as those who interleaved other tasks while programming. Participants utilized a variety of techniques to deal with high workload. Some chose to slow the aircraft down to “buy” time or shed or truncated a task, such as acknowledging an ATC traffic alert but then not personally scanning for the traffic. These two strategies tended to be used less often than others such as requesting vectors or alternate routing from ATC. In future studies, it would be informative to evaluate the use of strategies for management of high workload that are controlled by the pilot (e.g., slowing the aircraft, shedding tasks), as compared to those involving assistance from the outside (i.e., ATC). Both are certainly necessary and appropriate in various situations. We found that those who utilized methods they controlled, such as reducing airspeed, often accomplished the scripted tasks successfully.

The longstanding principle of completing as many tasks as possible during low workload periods to reduce the number that must be performed during periods of higher workload generally worked well for our participants, particularly the two who programmed the approach at KHSP very early. It would be interesting to examine in a future study the efficacy of this strategy for programming instrument approaches, even if it means that changes are required later. Task prioritization relative to the amount of time available is a critical part of workload management. Those participants who had not adequately prepared for the instrument approach at KHSP prior to the end of the lost pilot scenario were more likely than others to encounter difficulty in accomplishing the task successfully. Single pilots operating jets under NextGen must have a keen sense of the temporal aspects of flying tasks and use a variety of strategies to manage their workload to successfully complete their flights.

Single-pilot workload management is strongly associated with automation use and errors made when programming the automation. We found that when participants were confronted with high workload they tended to opt for a lower level of automation to reduce their workload in the moment (i.e., using autoflight modes as opposed to programming the G1000), even though that meant their overall ongoing workload might be greater. A mix of both input errors and other errors, which indicated a lack of understanding of how the automation and autoflight modes worked, were observed. Input error identification strategies developed for airline crews could be adapted for use by single pilots flying VLJs/ELJs (Berman, Dismukes, & Jobe, 2012). Targeted activities are needed during training to tease out pilot misperceptions and misunderstandings about how advanced automation functions.

Acknowledgments

This work was sponsored by FAA AFS-800, Flight Standards Service, General Aviation and Commercial Division, and funded through FAA ANG-C1, Human Factors Division. Early work on the study and scenario design was supported by NASA's Aviation Safety Program. We extend deep appreciation to our colleagues who reviewed earlier versions of this paper and to the Cessna Mustang pilots who participated in the study.

References

Berman, B.A., Dismukes, R.K., & Jobe, K.K. (2012). Performance data errors in air carrier operations: Causes and countermeasures. NASA Technical Memorandum (NASA TM-2012-216007) Moffett Field, CA: NASA Ames Research Center.

Burian, B. K. (2007). Very Light Jets in the National Airspace System. *Proceedings of the 14th International Symposium on Aviation Psychology*. Dayton, OH.

Burian, B. K., & Dismukes, R. K. (2007). Alone at 41,000 feet: Single-pilot operations in technically advanced aircraft. *Aero Safety World,* (11), 30-34.

Burian, B. K., Pruchnicki, S., Rogers, J., Christopher, B., Williams, K., Silverman, E., Drechsler, G., Mead, A., Hackworth, C., & Runnels, B. (2013). *Single pilot workload management in entry level jets.* Final Report. http:// humansystems.arc.nasa.gov/flightcognition/publications.html

Dismukes, K., Berman, B. A., & Loukopoulos, L. D. (2007). *The limits of expertise : rethinking pilot error and the causes of airline accidents*. Aldershot, Hampshire, England ; Burlington, VT: Ashgate.

Federal Aviation Administration [FAA]. (2008). *Airline transport pilot and aircraft type rating practical test standards for airplane*. FAA-S-8081-5F. Washington, DC: U.S. Government Printing Office.

Federal Aviation Administration. (2010). *Instrument rating practical test standards for airplane, helicopter, and powered lift*. FAA-S-8081-4E w/ changes 1 and 2. Washington, DC: U.S. Government Printing Office.

EXPLORING THE IMPACT OF ADS-B NEXTGEN TECHNOLOGY REQUIREMENTS FOR GENERAL AVIATION

Rachel Borsa
Purdue University
West Lafayette, Indiana
Professor Brian G. Dillman
Dept. of Aviation - Purdue University
West Lafayette, Indiana
Professor Timothy D. Ropp
Dept. of Aviation - Purdue University
West Lafayette, Indiana

The Next Generation Air Transportation System, or NextGen, represents the modernization of the United States air transportation system. It is a shift from legacy computer and radar based technologies to an integrated set of technologies, connected into a real time data sharing "system," shared by air traffic control, pilots and airports. This upgrade to our National Airspace System (NAS) involves potentially significant changes and financial considerations for many air operators. The upgrade and integration of one of these technologies, the Automatic Dependent Surveillance – Broadcast (ADS-B) transmitter, is among the mandatory upgrades being required by the Federal Aviation Administration within a few years. This paper explores the cost and benefits of this specific NextGen technology for general aviation owner/operators, takes a specific look at the implications for flight training schools, and details the many ways that general aviation pilots can access and use the new information in the cockpit.

Air transportation is a major economic engine for the United States, accounting for 5.2 percent of the United States Gross Domestic Product and contributing over $1 trillion annually to the national economy, and 11 million jobs (FAA, 2011b). The required growth and modernization of United States civil aviation is therefore paramount in order to sustain this benefit while meeting anticipated future demand for air travel. Growth and sustainability will be constrained if modern day issues related to safety, data sharing capabilities, aircraft noise, emissions, and fuel economy – as targeted by the NextGen initiative - are not addressed (US Government, 2003). The underlying nature of NextGen seeks to address issues of safety, capacity and efficiency which could jeopardize sustainable air transportation in the 21st century if these outcomes are not maintained as the NextGen system evolves. Without a major improvement to the current air transportation system, the United States risks falling behind in safety, efficiency and overall competitiveness within the global air transportation system.

Graduates from FAA approved flight programs must enter the industry familiar with the emerging technologies as well as the issues that are inherent within their utilization. This report explores background information for Automatic Dependent Surveillance – Broadcast (ADS-B), an onboard technology that will be required for aircraft including general aviation in most airspaces by the year 2020 (FAA-AC 90-114, 2011). This report looks at the implications for training schools with general aviation fleets and potential benefits of investing in and learning ADS-B technology early on in the deployment schedule for the technology. A purposeful early investment sooner rather than later (closer to an FAA mandated deadline) is believed to provide greater benefit to training programs and their students, the future NextGen professionals.

Background

The modernization of the United States National Airspace System (NAS) into the Next Generation Air Transportation System (NextGen) has been an ongoing process since its mandate by Congress in 2003 under the Vision 100 Aviation Reauthorization Act (US Government, 2003). This upgrade is being rolled out by the Federal Aviation Administration (FAA) pursuing convenient, dependable and hassle-free air travel that will also accommodate projected levels of increased demand for air travel worldwide in the coming decade (NextGen, 2011).

NextGen involves all levels of the aviation industry including, air traffic control, commercial air carriers, and general aviation. Premier among innovations on the flight operations side are new satellite and Global Positioning System (GPS) technologies in place of traditional ground based radar, and flight deck upgrades enabling more autonomous, real time fact-based decision making with the pilot in command of an aircraft in flight. As a result of this holistic 'systems' approach of shared data and more pilot autonomy (see Table 1), the industry is evolving from the traditional concept of air traffic control to a more relaxed concept of *air traffic management* (2011), although the term 'relaxed' may perhaps, in the view of the authors, be a misnomer.

There has been little significant change in air traffic control and airspace systems since the end of WWII, or the Jet Age (FAA Implementation, 2012). At one time the NextGen concept was relegated to "forward thinkers" looking to intangible future technology capabilities. However, with technology already demonstrating its capabilities and real deadlines quickly approaching, action must be taken. NextGen leverages both new technology as well as technologies that already exist (NextGen, 2011). For aircraft applications, the goal is to provide flight crews greater situational awareness and more control over what is happening in the aircraft, and to reduce time on the ground waiting for air traffic control instructions. Data acquisition and decision making traditionally assembled and relayed from the ground to the aircraft will occur much more substantially in the flight deck, with controllers assuming the role of traffic flow managers rather than ground controllers (FAA, 2011a).

Table 1.
NextGen is split into six different components (FAA, 2011a). They are:

- Automatic Dependent Surveillance-Broadcast (ADS-B) (the focus of this report)
- System Wide Information Management
- NextGen Network Enabled Weather
- NAS Voice System
- Collaborative Air Traffic Management Technologies, and
- DataComm Data Communications.

There are several different aims that the FAA has for NextGen. The FAA wants to keep delays low and give the pilots more flexibility to maneuver around weather problems. In addition, the system is being designed to make a smaller carbon footprint with more precise flights, to establish "one seamless sky," create more jobs in more cities, and to keep travelers safe (FAA, 2011).

The Role of ADS-B

ADS-B continues to play a prominent role in NextGen and is one of the flagship technologies highlighted by the FAA. By the year 2020, the FAA will be implementing the rule that all aircraft must be equipped with onboard position reporting and identification technology called ADS-B (FAA-AC 90-114, 2011; FAA Final Rule, 2010). In September of 2010, the FAA approved satellite-based technology for air traffic controllers in areas with the coverage (FAA, 2012). Although planning has been underway for nearly a decade, the pending requirement for ADS-B remains problematic, especially for general aviation operators. Questions range from "*What is it and how do I use it?* to *"How much does it cost and what's the real benefit?"*

ADS-B System Description. Each word in Automatic Dependent Surveillance-Broadcast is descriptive of the overall system (ADS-B Technologies, 2011):

Automatic: always on and requires no operator intervention.

Dependent: system is dependent on an accurate Global Navigation Satellite System (GNSS) signal (also known as Global Positioning System GPS),

Surveillance: the method used for determining position of aircraft, vehicles or other assets.

Broadcast: continuously broadcasts aircraft position and other data to any ADS-B equipped aircraft, or ground station equipped to receive ADS-B.

ADS-B Out provides information similar to what the Transponder/Decoder system provides (secondary radar), but the way information is received is completely new. Ground stations and aircraft both require ADS-B Out for this technology to work. ADS-B Out will be the requirement that the FAA will mandate. Aircraft receive position data from a GPS satellite constellation in space and then simultaneously broadcast their position and all other data to any aircraft equipped with ADS-B In technology and ground station equipped to receive it (ADS-B Technologies, 2011).

As stated earlier, the only ADS-B standard required by 2020 for all aircraft is to have ADS-B Out installed for flight in ADS-B airspace (FAA-AC 90-114, 2011). Specifically, ADS-B Out will be required in (FAA, 2010a):

- Class A, B, and C airspace
- Class E airspace within the 48 contiguous states and District of Columbia at and above 10,000 feet MSL, excluding the airspace at and below 2500 feet above the surface
- Class E airspace at and above 3000 feet MSL over the Gulf of Mexico from the coastline of the US out to 12 nautical miles
- Around those airports identified in 14 CFR part 91, Appendix D

The ADS-B Out rule does not apply to an aircraft that was not originally certificated with an electrical system or that has not been certified with such a system installed, including balloons and gliders (FAA, 2010b).

It would appear that simple. Radar could degrade with range, atmospheric conditions, or target altitude. Also, update intervals do not depend on the rotational speed or reliability of mechanical antennas. ADS-B will be faster than radar and more fail-resistant. When ADS-B Out technologies are paired with ADS-B In technologies in the aircraft, the result will be information that can be utilized with almost complete situational awareness. ADS-B Out can tell us aircraft speed, position, direction, relative altitude, and aircraft tail numbers (Dillman, 2012). There will be virtually no more blind spots and if utilized correctly, pilots will be much safer and more aware in the air.

ADS-B Out can be equipped one of two ways. The owner can install a 1090 megahertz (MHz) extended squitter broadcast link, or the Universal Access Transceiver (UAT) broadcast link (2010). The extended squitter and the UAT are both certified by Technical Standard Orders, TSO-C166b and TSO-C154C, respectively (2010). Purdue University acquired a fleet of Cirrus SR-20's in 2010. With those airplanes, the University bought the GTX 33ES transponder (Dillman). The ES stands for extended squitter and Purdue bought that transponder because it already had ADS-B Out capability and it is currently being utilized in the aircraft.

A steep learning curve remains for pilots utilizing ADS-B to fully realize the potential benefits available. While ADS-B Out will be required by the FAA, the true benefit for pilots operating within the system will be the information provided by ADS-B In. ADS-B In will contain Flight Information Services-Broadcast (FIS-B) and Traffic Information Services- Broadcast (TIS-B) (2010). This will be an excellent resource for pilots because certain aspects of the information available will be free, and pilots can access this information from in the cockpit (2010). If the weather begins to deteriorate, the pilot can scroll across her screen and find where the weather is good enough to land, or the pilot can fly above or around the weather. Some other benefits of having ADS-B In are: enhanced visual acquisition, enhanced visual approaches, final approach and runway occupancy awareness, airport surface situational awareness, conflict detection, merging and spacing, Cockpit Display of Traffic Information (CDTI), and Cockpit Assisted Visual Separation (CAVS) (2010). This is where having a compatible display, like the glass cockpit design of the Cirrus SR-20's, will be required to fully benefit from the technology. The pilot will need a big enough screen so that she can see all of the information provided. An alternative platform is the incorporation of tablet technologies such as the Apple IPad. For a somewhat small investment, the ADS-B In information can be accessed with relative ease. The options for compliance of the ADS-B Out mandate are widespread and the available access to the ADS-B In information is available from a variety of resources as well. Because of this fact, it is imperative that the pilots operating a given aircraft be familiar with the resources available and, more importantly, how they will access the desired information.

Viewing ADS-B as an Educational Tool

NextGen impacts General Aviation on several levels. General aviation training and operators will need to do more research about ADS-B technology than the average airline pilot. GA pilots will, in a sense be more in charge of deciding what device they receive their information on, where the original information comes from, understanding the different limitations, and when and where to best use the information. Owner/operators have expense and transitional training to consider and rightfully so. Similarly, from a GA training fleet perspective, while initial cost structure can be high, the positive outcome is a better equipped student.

As an example, Purdue's fleet of Cirrus SR-20s represents an excellent platform for the new ADS-B technology. While the cost of this fleet ultimately adds to student fee structure, the student also is introduced to NextGen style technologies from the beginning of training. The Cirrus aircraft are glass cockpit and contain two screens; the Primary Flight Display (PFD) and the Multi-Function Display (MFD). The PFD has all of the information that traditional round style instrumentation provide on older aircraft. The MFD is where the majority of the information pertinent to ADS-B is located. The Traffic Avoidance System (TAS) shows the position of other airplanes with transponders and ADS-B Out equipment. The ADS-B allows pilots to see the direction of flight of conflicting traffic, allows the determination if the aircraft is ascending or descending, and can identify the tail number of the other ADS-B aircraft. Seeing all of that additional information only occurs when the planes have ADS-B Out. ADS-B is a useful tool in facilitating pilots to see and understand more of what is going on around the pilot. When utilized correctly, situational awareness increases exponentially.

Again, it is recognized that the cost of a fully equipped ADS-B can be steep for the individual aircraft owner/operator. Market price at the time of this report for ADS-B capability is $8395US (Garmin, 2012). Then, the owner has to pay for the maintenance to install the transponder with ADS-B capability; which is usually an extra $1000. Just ADS-B Out can cost anywhere from $2000 to upwards of $10,000. However, there are other options and alternate routes that owners can take to create a cheaper purchase. The FAA is currently providing benefits and price cuts to those who install the equipment now. Also, the FAA is more lenient on what way the owner installs the ADS-B capable equipment. It can be in the form of the transponder or in the form of an antenna, but if it is on the plane, then it counts. In the future, the FAA could become stricter about what equipment can and cannot be used to transmit ADS-B Out.

ADS-B In is received in a different way than ADS-B Out. Purdue has a GTS 800 which is capable of receiving TIS-B and FIS-B when that becomes functional to pilots (Garmin, 2012). With the MFD, the pilot can scroll to different screens and she will see weather on one page, including METARs, TAFs, Sigmets, Airmets, TFRs, and Radar (Lawrence, 2011). In conjunction with the TAS system, the ADS-B allows the pilot can see all of the traffic in the area. ADS-B In can "see" an area of 100 nm (2011). The most common distance to set the distance is eight miles. When ADS-B In locates an aircraft within those eight miles, or however the pilot changes the settings, a warning will sound; saying, for example, "Traffic, two 'o'clock, seven point five miles, three hundred feet below" (2011). This feature makes the pilot's job easier by finding the potentially harmful traffic and avoiding it. ADS-B In really is a remarkable piece of technology for those that have the ability to access the information.

There are many different ways to illustrate that ADS-B technology is worth the money, time, and practice, especially at a training school like Purdue. Other than verifying that the aircraft you're operating is equipped with ADS-B Out and an awareness of which airspace it is required to operate, there is nothing really to learn involving ADS-B Out. It is an automatic signal and cannot be controlled by the pilot. However ADS-B In, once it is functional, will be an invaluable resource for the student aviators at Purdue and something that will need to be incorporated into the plethora of other information sources that are available. Unlike aircraft from decades ago, there are virtually unlimited possibilities for obtaining source material to complete the pilot's responsibilities. They will be seeing and reading the same information that airline pilots utilize in flight. The technology should be streamlined across the industry so that every system is somewhat alike which will allow for seamless transitions between various aircraft and operators. Other data sources include constant weather updates on days that are marginal VFR, enabling more precise predictions and flight plan decision making including weather avoidance and go/no-go scenarios.

Conclusion

Regardless of cost concern or learning curves, NextGen technology integration continues, and it is time to move from the questions of "why and when" to the next question of "how". ADS-B In offers a consistent format of information, albeit through varied mechanisms and systems. As the NAS evolves, opportunity exists at the educational level to shape novice aviators into better systems thinkers by learning the requisite methods and approaches necessary to interact with a complex air transportation system before they enter it. This in turn will facilitate a more informed human operator and achieve crucial future safety outcomes.

At the university level, ADS-B is seen as a significant enabler for sharpening 21st century flight deck command, awareness and decision-making skills. This in turn prepares students to assimilate into airline careers already being transformed by NextGen. On top of all that, proper utilization of ADS-B makes the skies safer. Soon we will not think of ADS-B as an expensive luxury, but as a necessary human-in-the-loop technology in the flight deck.

# #

References

ADS-B Technologies (2011). ADS-B technologies. Retrieved from http://www.ads-b.com/

ADS-B Technologies (2011b). ADS-B Link Augmentation System (ALAS) Retrieved from http://www.ads-b.com/news.htm.

Dillman, B. (2012, February). Interview by R. Borsa [Personal Interview]. Ads-b.

FAA (2010a). Automatic dependent surveillance-broadcast (ads-b) out performance requirements to support air traffic control (atc) service. **Federal Register** / Vol. 72, No. 193 / Friday, October 5, 2007 / Proposed Rules (FAA-2007-29305). P. 96958.

FAA (2010b). Surveillance and broadcast services. Retrieved from http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/enroute/surveillance_broadcast/general_information/

#

FAA (2011a). FAA NextGen What's New. Retrieved from http://www.faa.gov/nextgen/today/whatsnew/

FAA (2011b). FAA NextGen implementation Plan - 2011. Introduction, p. 8.

FAA (2012). FAA NextGen implementation plan - March 2012. Retrieved from website: http://www.faa.gov/nextgen/slides/?slide=3

FAA-AC 90-114, (2011). Advisory Circular 90-114 Automatic Dependent Surveillance-Broadcast (ADS-B) Operations. Par 1-1, page 1. Guidance for complying with 14CFR Part 91.225/227. Automatic Dependent Surveillance-Broadcast (ADS-B) Out performance when operating in designated classes of airspace within the UNITED STATES National Airspace System (NAS) after January 1, 2020. December 8 2011.

FAA Final Rule (2010), FAA final rule on ads-b. (2010, May 27). Retrieved from http://www.avweb.com/avwebflash/news/FAA_FinalRule_ADSB_TextExcerpt_202632-1.html

Garmin. (2012). Gtx 33/33d. Retrieved from http://www8.garmin.com/products/gtx33/

Lawrence, J. (2011). Skyvision xtreme. *ADS-B, here and now*, Retrieved from http://www.planeandpilotmag.com/products/tech-talk/skyvision-xtreme.html

BEYOND MULTITASKING: HUMAN FACTORS IMPLICATIONS FOR SINGLE-PILOT OPERATIONS IN THE NEXTGEN ENVIRONMENT

Nestor Fenoy Bermudez
Florida Institute of Technology
Melbourne, FL
Donna F. Wilt, Ph.D.
Florida Institute of Technology
Melbourne, FL
John E. Deaton, Ph.D.
Florida Institute of Technology
Melbourne, FL

Operations within the Next Generation Air Transportation System (NextGen) will be a source of new complexities and unique challenges for the single pilot. Human factors research and development will be required to ensure safe single-pilot operations. New avionics, the increase in procedural complexity, and the elevated standards for required navigation performance are the source of many issues that need to be identified before they become a threat to safety. This paper presents a taxonomy developed to identify the potential human factors issues that could impact single pilots flying within NextGen. A series of NextGen capabilities, scenarios, and Automatic Dependent Surveillance-Broadcast (ADS-B) applications were extracted from the literature and examined to determine potential single-pilot human factors issues. A priority list was created to emphasize the issues that constitute the biggest threat to single pilots. The implications of the results are discussed in this paper, and recommendations are given for future research.

The Federal Aviation Administration (FAA) has defined the Next Generation Air Transportation System (NextGen) as a set of technologies, and actions enabled by those technologies, that represents the most significant overhaul of the National Airspace System (NAS) in U.S. history (Federal Aviation Administration [FAA], 2011). This overhaul consists of the transformation of the air traffic control system from a ground-based system to a satellite-based system. A satellite-based system allows for the use of Global Positioning System (GPS) technology to shorten routes, reduce traffic delays, increase airspace capacity, and save time and fuel (FAA, 2007). Operations within NextGen will be a source of new complexities and unique challenges for the single pilot. The introduction of new avionics, the increase in procedural complexity, and the elevated standards for required navigation performance (FAA, 2011) will generate a series of issues that will need to be identified before they become a threat to operational safety for the single pilot. These issues will be system wide, for they will affect both pilots and air traffic controllers. However, the lower amount of resources available to single pilots, in comparison to larger crews, makes them a research priority.

The single pilot is already responsible for many different tasks and procedures, often concurrent, in the current NAS. These tasks can range from visually scanning out the window, monitoring the instruments, and gathering information from the displays, to processing all of the received information, using it to make decisions, and ultimately flying the airplane in a safe manner (FAA, 2008). Therefore, it is crucial to determine any potential increase in such task demands that NextGen will impose on the single pilot in order to avoid potential overload.

Previous research (FAA, 2009b) suggests that mental workload will play a very important role in the success or failure of the single pilot when operating in the NextGen environment. Furthermore, the single pilot's ability to achieve and maintain a safe level of situation awareness (SA) will also be heavily dependent on the level of workload required to process the information made available in the cockpit. Thus, this presents a research gap that requires further study in order to determine the best solution, or approach, for any potential human factors issues that may arise in this area.

Methodology

This research supports the previous work done by the FAA (FAA, 2009b). The research goal of this project was to develop a complete list of significant human factors issues and challenges that represented the main potential problems with single-pilot operations in NextGen. This goal was divided into three steps: a) a literature review of

NextGen-related documents, b) the examination of specific human factors research that addressed human factors issues in multi-person crews in NextGen, and extracting single-pilot related information, and c) the development of a taxonomy that related human factors issues to several NextGen areas. As a result of this project, recommendations were presented for the focus of future research studies, which included the particular single pilot human factors issues that could represent a threat for single-pilot operations.

NextGen Single-Pilot Items

Several Federal Aviation Administration (FAA) and Joint Planning and Development Office (JPDO) documents were reviewed that included NextGen scenarios, capabilities and Automatic Dependent Surveillance-Broadcast (ADS-B) applications (JPDO, 2007; FAA, 2009a; FAA, 2010a; FAA, 2010b). Overall, 24 scenarios, capabilities and applications were initially extracted from the NextGen documents. For the sake of simplicity, the NextGen scenarios, capabilities and ADS-B applications were combined into a broader category named NextGen Single-Pilot Items. After the Subject Matter Experts (SMEs) reviewed the 24 items, and considered their potential impact on the single pilot, the list was reduced to 18. The rationale for the elimination of some of these items was two-fold. On the one hand, some of the items were found by the SMEs to have many similarities, which would have created redundancy in the results. On the other hand, the SMEs believed that the human factors issues generated in some items would not be unique to the single pilot; but rather independent of the size of the crew, thus making them irrelevant for this research. As shown in Table 1, the remaining 18 items were then categorized according to the phase of flight (i.e., departure, enroute, arrival, approach, surface, or overlaying all phases).

Table 1.
Finalized List of 18 NextGen Single-Pilot Items Extracted from the NextGen Literature and Organized by Phase of Flight.

Departure	
	RNAV/RNP Departure Procedures
	Equivalent Visual Operations for Departure
Enroute	
	Flight-Deck Based Interval Management-Spacing
	Delegated Responsibility for Horizontal Separation
	Flight-Deck Based Interval Management with Delegated Separation
	Independent Closely Spaced Routes
	Time-Based Metering Using RNAV and RNP Route Assignments
Arrival	
	Point in Space Metering
Approach	
	Paired Closely Spaced Parallel Approaches
	Traffic Situation Awareness for Visual Approach
	Improved Operations to Closely Spaced Parallel Runways
	Use Optimized Profile Descent
	Equivalent Visual Operations for Approach
Surface	
	Enhanced Surface Traffic Operations
	Airport Traffic Situation Awareness
Overlay All Phases	
	ADS-B Integrated Collision Avoidance
	Use of Data Communications
	Network-Enabled Information Access / Aeronautical Information Services

Note. NextGen Single-Pilot Items consist of NextGen scenarios, capabilities, and ADS-B applications.

Single-Pilot Human Factors Taxonomy and Issues

The second step in the development of the list of human factors issues was examination of work by Funk, Mauro, and Barshi (n.d.) of a research study that addressed human factors issues in multi-person crews in NextGen. In their final report, Funk et al. (n.d.), included a database with 236 specific human factors issues, organized into 10 categories, that could potentially affect any type of crew in a NextGen environment. Due to the narrower focus of this single-pilot study, this step extracted a set of human factors issues tailored specifically to single-pilot operations.

The extraction process consisted of determining which human factors issues a) were not exclusively related to crews with more than one person, b) maintained a relationship to single-pilot operations, c) were not so similar to others that would cause the results to be duplicative, and d) were measurable by a previously developed and validated scale. A total of 23 NextGen human factors issues were extracted from Funk et al. and organized into 6 broader categories (see Table 2). The extracted issues became the foundation for a taxonomy that further linked single-pilot human factors issues to the previously established 18 NextGen Single-Pilot Items.

The third and final step of the analytical process was a more detailed examination the human factors issues by two SMEs. This detailed examination determined which of the human factors issues in the list (see Table 2) could potentially manifest itself within the different NextGen Single-Pilot Items. A series of comments and questions regarding the impact of each of the human factors issues on the NextGen Single-Pilot Items were developed.

Table 2.
List of 23 Human Factors Issues Related to NextGen Single Pilots.

Pilot Characteristics		Controls Interaction	
	Aeronautical Decision Making		Feedback
	Single Pilot Resource Management		Manual Skill
	Situation Awareness	System Interface	
	Allocation of Attention		Representation & Processing
	Risk Assessment		Interface Functionality (System Access)
	Memory		Complexity
	Other Mental Workload		Cues & Alarms
	Physical Workload		Delay
	Stress	ATM	
	Fatigue		Communication & Collaboration
	Error Management	Other	
Automation Interaction			Training
	Complacency		
	Monitoring		
	Managing Automation Failure		

Results

The results obtained from the comparison of the NextGen Single-Pilot Items with the individual human factors issues were synthesized into a taxonomy table (see Figure 1). In this taxonomy, human factors issues are matched to those NextGen Single-Pilot Items in which they have an impact for the single pilot. Using the taxonomy, researchers identified the most prevalent issues across the different NextGen Single-Pilot Items (see Table 3). Then conclusions were extracted regarding the criticality of specific human factors issues and recommendations were made regarding which of these issues should have research priority above the others.

		Pilot Characteristics											Automation Interaction			Controls Interaction		System Interface					ATM	Other
	NextGen OV-6c Scenarios (S), Capabilities (C) & Applications (A)	Aeronautical Decision Making	SP Resource Management	Situation Awareness	Allocation of Attention	Risk Assessment	Memory	Other Mental Workload	Physical Workload	Stress	Fatigue	Error Management	Complacency	Monitoring	Managing Automation Failure	Feedback	Manual Skill	Representation & Processing	Interface Functionality (System Access)	Complexity	Cues & Alarms	Delay	Communication & Collaboration	Training
	Departure																							
1	RNAV/RNP Departure Procedures (Existing)	x		x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x
2	Equivalent Visual Operations (EVO) for Departure (C)	x	x	x	x	x			x	x		x	x	x	x	x		x	x	x	x	x		x
	Enroute																							
3	Flight-Deck Based Interval Management-Spacing (FIM-S) (A 6)	x		x	x		x	x		x	x	x	x	x	x	x		x	x		x		x	x
4	Delegated Responsibility for Horizontal Separation (S 13)			x	x			x					x	x	x	x		x	x		x		x	x
5	Flight-Deck Based Interval Management with Delegated Separation (FIM-DS) (A 8)	x		x	x		x	x		x	x	x	x	x	x	x		x	x		x		x	x
6	Independent Closely Spaced Routes (A 9)			x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x
7	Time-Based Metering Using RNAV and RNP Route Assignments (S 7 & 8)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	Arrival																							
8	Point In Space Metering (S 7)	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x
	Approach																							
9	Paired Closely Spaced Parallel Approaches (A 10)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
10	Traffic Situation Awareness for Visual Approach (A 2)	x		x	x		x	x	x	x	x	x	x	x	x	x		x	x	x	x		x	x
11	Improved Operations to Closely Spaced Parallel Runways (CSPR) (S 15)	x	x	x	x	x		x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x
12	Use Optimized Profile Descent (OPD) (S 12)	x	x	x	x			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
13	Equivalent Visual Operations (EVO) for Approach (C)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	Surface																							
14	Enhanced Surface Traffic Operations (S 6)	x	x	x	x	x		x	x			x		x	x	x		x	x		x	x	x	x
15	Airport Traffic Situation Awareness (A 3)	x	x	x	x	x	x	x				x	x	x		x		x	x	x	x	x		x
	Overlay All Phases																							
16	ADS-B Integrated Collision Avoidance (A 15)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
17	Use of Data Communications (C)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
18	Network-Enabled Information Access / Aeronautical Information Services (C)	x	x	x	x	x	x	x	x	x	x	x	x	x	x			x	x	x	x	x	x	x

Figure 1. Taxonomy table of NextGen Single-Pilot Items vs. single-pilot human factors issues.

Conclusion and Recommendation

This research supports that additional research is needed to allow the single pilot to safely fly in the NextGen environment. Future research will need to identify a) factors that could affect the single pilot's situation awareness, b) guidelines that could aid the pilot in prioritizing tasks under high levels of stress, c) requirements for cues and alarms that can inform the pilot about deviations or procedure errors, and d) best methods for presenting the necessary information and interfacing with the avionics in a way that maintains a reasonable level of physical and mental workload while avoiding complacency.

Future avionics design should consider the single pilot's challenges in monitoring and interfacing with instruments and displays inside the cockpit while also maintaining a visual scan outside the cockpit. Proper cues, alarms, and pre-processing will be essential to offload some of the monitoring and workload requirements, and to mitigate issues with allocation of attention that NextGen applications will impose on the single pilot. Furthermore, new information needed by pilots as a result of NextGen applications needs to be presented in a clear and concise manner. The single pilot should be able to perceive the needed information, easily interpret it, and use it to make projections into the future.

Thus, research goals for future studies are a) the balance between clarity and conciseness within new NextGen displays; and b) interface functionality (i.e., system access). The single pilot must be able to navigate through the different interfaces and access whatever information is needed in a way that is not only easy to learn and remember, but also time efficient. Considering all of the other tasks that must be completed by this one person, no time can be wasted trying to locate the required information needed at any point in time. This could also impact the single pilot's level of situation awareness in that, if too much time is wasted trying to obtain information, not enough will be available for the perception and further processing of that information.

Table 3.
Prioritized List of 23 Human Factors Issues Based on Their Number of Appearances Across the 18 NextGen Single-Pilot Items.

Human Factors Issues	Number of Appearances
Situation Awareness	18
Allocation of Attention	18
Monitoring	18
Representation & Processing	18
Interface Functionality (System Access)	18
Cues & Alarms	18
Training	18
Other Mental Workload	17
Complacency	17
Managing Automation Failure	17
Feedback	17
Aeronautical Decision Making	16
Error Management	16
Communication & Collaboration	16
Physical Workload	14
Stress	14
Complexity	14
Delay	14
Risk Assessment	13
Memory	13
Fatigue	13
Single Pilot Resource Management	12
Manual Skill	9

Note. No prioritization exists between those human factors issues with the same number of appearances.

Acknowledgements

This research was sponsored by the Federal Aviation Administration as part of a grant awarded to the Florida Institute of Technology titled “Determination of Human Factors Issues and Recommended R&D Requirements for Single-Pilot Aircraft Operations in the NextGen Environment.”

References

Federal Aviation Administration. (2007). *Fact Sheet: NextGen*. Retrieved from http://www.faa.gov/news/fact_sheets/news_story.cfm?newsid=8145

Federal Aviation Administration. (2008). *Pilot’s Handbook of Aeronautical Knowledge*. Retrieved from http://www.faa.gov/library/manuals/aviation/pilot_handbook/

Federal Aviation Administration. (2009a). *National Airspace System Enterprise Architecture Framework (NASEA), Operational Event-Trace Description (OV-6c), Version 0.1*.

Federal Aviation Administration. (2009b). *NextGen Human Factors in Single Pilot Operations, Version 5.1*.

Federal Aviation Administration. (2010a). *Application Integrated Work Plan, Version 2.0*.

Federal Aviation Administration. (2010b). *NextGen Mid-Term Concept of Operations for the National Airspace System, Version 2.0*.

Federal Aviation Administration. (2011). *AVS Work Plan for NextGen 2011*.

Funk, K., Mauro, R., & Barshi, I. (n.d.). *Nextgen flight deck human factors issues, final report*. Unpublished manuscript.

Joint Planning and Development Office. (2007). *Concept of Operations for the Next Generation Air Transportation System, Version 2.0*.

DOES SUPPLEMENTARY COMPUTER GENERATED CUEING ENHANCE CONTROLLER EFFICIENCY IN A CONGESTED COMMUNICATION ENVIRONMENT?

Laurienne C. R. A. Santana, Brazilian AF; Brent T. Langhals, Lt Col, USAF; Michael E. Miller, PhD
Air Force Institute of Technology
Dayton, Ohio

Vic Finomore, PhD
711th Human Performance Wing, WPAFB, USAF
Dayton, Ohio

Air traffic controllers are often required to simultaneously communicate with several aircraft over multiple radio frequencies. As a result, during peak loading, it is common for the controller to receive multiple concurrent communications, each from a different aircraft, making it difficult to discern all necessary auditory communications. To address this problem, a modified air traffic control (ATC) interface was prototyped with the goal of increasing controller-to-pilot communication efficiency. This prototype involved an automated text to speech system which displayed supplementary text in an on-screen text box, echoing aircraft transmissions in the event of an indiscernible radio call. The prototype was then evaluated by a group of 35 participants, all with ATC experience at the Air Force controller's school house Keesler AFB, MS. Results of the study indicated text cueing may be useful for improving controller comprehension of pilots' transmissions.

In the future, ATC will be challenged by the expected increase in the amount and complexity of air traffic (Stein & Garland, 1993). The Federal Aviation Administration (FAA) Aerospace Forecast stated that in 2011, total activity at FAA en-route centers increased by 1.8 percent from the previous year to a total of 41.2 million, the fastest growth since 2005. This report also projected the number of aircraft handled at FAA en-route centers, which separate high altitude traffic, to increase by 50 percent (FAA, 2012). As the air traffic load increases, similar increases in communication load expected to occur. Frequency congestion is a sensitive situation that occurs during the communication process between pilots and controllers and "has become a factor that severely constrains the capacity of the National Airspace System (NAS)" (Rantanen et al, 2004).

Controllers often deal with a congested communication environment and are frequently required to communicate with several aircraft over multiple radio frequencies. As a result, during peak loading, it is common for the controller to receive multiple concurrent communications, each from a different aircraft on a different radio frequency, challenging controllers' ability discern audio messages received from multiple pilots simultaneously (this phenomenon is known as "step-on"). As controllers monitor all frequencies through a single earpiece, it is impossible for the controller to interpret concurrent communications, resulting in additional workload as the controller must then decipher the source of each communication and request each aircraft to repeat the communication asynchronously. In some extreme cases, improper or misunderstood communication has contributed to accidents (Rantanen et al, 2004).

This study investigated the communication between pilots and controllers, which is the foundation of air traffic control (Hopkin, 1995). Specifically a method is proposed to improve the existing air traffic controller (ATC) workstation by providing text transcriptions of pilot audio communications to controllers. The text allows the controller to respond without having to transmit a request for a repeat, wait for a re-transmission and then respond. This cumbersome process to overcome "step on" adds exponentially to operator workload and frequency congestion. The proposed system design provides the controllers with accurate transcription of pilot audio communications in a usable and readable format, located in a single window and displayed in a sequential manner. Such a system is expected to decrease the time and steps needed to answer pilots' requests, increasing controller's efficiency and safety within the National Airspace System (NAS).

Background

ATC is a dynamic environment where controllers constantly receive a large volume of information from multiple sources to monitor changes in the environment, make decisions, and perform effective actions in a timely manner (Xing & Manning, 2005). ATC system intricacies include thousands of separate facilities all communicating with each other and handling information via different sources (e.g., radar screens, paper or electronic flight progress strips, radio and interphone communication). Air traffic controllers have to deal with these different sources of information and control complex, dynamic and time-constrained traffic situations to identify and resolve potential conflicts and risky relationships between aircraft. Controllers have to perceive, analyze, comprehend, and anticipate multiple characteristics and flight paths of many aircraft. Incoming and departing aircraft as well as constantly changing aircraft trajectories combine to create new traffic relationships for evaluation. All of this must be performed in real-time and for the purpose of separating aircraft and issuing safety alerts (FAA, 2011).

Audio Comprehension and Task Difficulty

To provide systems and cues that better support the decision making process and enhance situation awareness it is important to understand how humans process information. Wickens' multiple resource theory (MRT) identified the structural dimensions of human information processing and proposes that there are four dimensions (processing stages, perceptual modalities, visual channels, and processing codes) with two categorical levels of each dimension (i.e. separate resources) (2008). This theory suggests that cognitive resources used to process information from one modality compete less with the resources used to process input from another modality than resources used to process input from the same modality. The availability of these resource channels results in a task-sharing benefit if information is distributed across different resources. According to this theory, an improvement in performance would be expected when different input modalities are addressed in a multitask environment (Parasuraman & Metzger, 2006).

Given the potential for decreased understanding arising from simultaneous speakers, targeting information to a second channel, according to the MRT Model, would be useful to explore. Specifically, text cueing may be particularly useful, in the sense that it adds another modality to improve recognition of audio transmissions and provides temporal stability, providing the operator access to each of the communications in the event that simultaneous verbal communications are broadcast over multiple frequencies.

H1: Supplementary text transcription of pilots' communication will increase controllers' comprehension of pilots' transmissions.

Furthermore, text cueing may also improve ATC task difficulty in the sense that it will reduce the need for say-again transmissions as the controllers would possess a secondary, temporally stable means to comprehend pilot's transmissions. As the controllers will not need to follow a cumbersome and time consuming process to request communication retransmission, text cueing should reduce the perceived difficulty of communication tasks.

H2: Text-cueing will reduce perceived difficulty of the communication task.

Memory Recall

Short-term memory accuracy has been a fundamental talent required by controllers to facilitate decisions and actions to ensure safety. According to Tindall-Ford and colleagues, the use of dual-mode presentation techniques (e.g., auditory text and visual diagrams) with students can result in superior learning as compared to equivalent, single-modality formats (Tindall-Ford, et al., 1997). The authors state this modality effect may be attributed to an effective expansion of working memory.

With the use of text transcription on screen, information presentation is shared between auditory and visual methods, in an attempt to control overuse of either mode, which is desired for the design of ATC systems (Cardosi & Murphy, 1995). When controllers listen and read pilots transmissions it is expected the presentation of the redundant verbal information to multiple modalities will improve memory recall. As a result, text cueing should improve controller's recall of what was seen and heard within the air traffic scene.

H3: Supplementary text-cueing will improve memory recall of pilot transmissions.

Distraction

While it is expected that text cueing would have positive impacts to ATC, care must be taken to understand any negative impacts to controller performance and ultimately flight safety. Distraction is a psychological error mechanism that is known to affect performance (Shorrock & Kirwan, 2002). In their paper describing methods for analysis of cognitive errors in ATC, the authors stated that distraction can affect different cognitive domains, such as: perception and vigilance (i.e. errors in visual detection and visual search, and errors in listening); memory (i.e. forgetting or mis-recalling temporary or longer-term information, forgetting previous actions, and forgetting planned actions); and action execution (i.e. actions or speech performed not as- planned). For this reason, potential causes of distraction in ATC should be carefully considered in the design of new aids to ensure that controllers' performance will not be impaired.

Today controllers are not experienced with supplementary text displayed on the side of the radar screen, but focus their visual attention on the radar display and other safety critical visual information. Ensuring safety is a priority for the design of ATCS. Although the text box was carefully designed to reduce distraction such that safety would not be compromised as the controller was forced to move their eyes from the radar display to the text box, it is possible that adding additional features to an already busy ATC workstation could increase controller distraction.

H4: Supplemental text box will increase controller´s distraction.

Methodology

Participants

The sample population involved thirty-five military ATC personnel, including 23 instructors and 12 students at a training installation at Kessler Air Force Base, MS. The students each had a minimum of 2 months of training prior to the experiment. Participation was voluntary and data was collected during a one week period. All of the participants were familiar with the simulated air space and the software used in the experiment.

Experiment Design

The experiment employed as a within subjects design where participants were exposed to two, three minute scenarios: control and experimental. The control scenario consisted of a standard radar screen, familiar to controllers, to use as a baseline to compare performance and did not include any text boxes. The experimental scenario had the text box containing a transcription of audio communication. A brief training scenario was presented to each participant before the scenario to explain how the text should appear on the screen. All of the scenarios used for this study were previously recorded and controllers were not permitted to alter them to standardize scenario experiences and allow comparison of subjects' performance. The scenarios were designed to be as realistic as possible to test the hypotheses of interest. Also, the control and experimental scenarios were developed to be equally complex to provide reliable comparison between controllers' performance regarding just the addition of the technology.

SIGNAL Air Traffic Control Simulation Program software, provided by the FAA, was used to develop and record the scenarios. The Multi-Modal Communication (MMC) System (Finomore et al, 2011) software was used to provide speech-to-text transcription. For the purpose of this study, a 100% accurate transcription of the audio communication was assumed since the objective of the study was to collect initial data of controllers' performance and perceptions of the advantages of the additional text, and not to test the technology itself. Text display was delayed by a period of time consistent with the recognition time required by the MMC system.

Measurements

Participants were required to monitor the display as they would in a regular air traffic control task; however they could only interact with the fictitious scenario pilots by clicking the mouse one time if they felt that a "SAY AGAIN" transmission was necessary. The number of mouse clicks was applied to test the first hypothesis. After the controllers were exposed to the air traffic scenario, a post-study questionnaire was administered regarding what they saw and heard during the experiment. To test the second and fourth hypotheses, participants' self-reported feelings of perceived difficulty and distraction were applied. For the third hypothesis, a query technique (Adams et al, 1995) was used to measure recall of specific pilot transmission requests. The post-test questionnaire also asked the participants to assess the realism of the scenarios and to indicate whether they found text cueing useful.

Results and Discussion

The two scenarios were perceived to be equally realistic. Also, 33 of the 35 participants reported they found the text cues useful in the performance of their task.

The primary dependent variables for this study were number of clicks (H1), self-reported difficulty (H2), percent correct of recall answers (H3), and self-reported distraction (H4). The main independent variable was scenario. Demographic and other background data was collected through a pre-study questionnaire and were considered as potential covariates during the analysis in order to explore their potential influence on controllers' performance. A t-test was run on the data with a 95% confidence interval to determine statistical differences between experimental and control scenarios for each hypothesis. In the cases that the t-test showed significance, a stepwise regression was conducted to determine if other independent variables may have also influenced the dependent variable.

For H1, t-test indicated that number of clicks for the experimental group, mean with a standard error of 2.34, were significantly lower than for the control group, mean(2.34 with a standard error of 1.327(t(68) = 2.638, p < .010). Therefore, stepwise regression was run and the model indicated an adjusted R^2 of .477 and included the additional independent variables of open mind, familiarity, multitask and experience ($F(69) = 13.609$, $p < .01$). Participants clicked the mouse less frequently when they reported they believe that new technologies can improve ATC, ; are familiar with the text on-screen technology; are able to multitask and feel comfortable listening and reading at the same time and had less experience in ATC. Individuals with more experience clicked more frequently and may have been more resistant to the change in presentation as it was it significantly different from their standard workstation.

The findings for H1 indicate that the experimental group had an increase in comprehension, requiring fewer interactions for communication clarification. The reduced number of transmissions would also potentially reduce frequency congestion. Another important contribution of the reduced number of transmissions is that it inherently reduces time needed to answer each pilot´s transmission, allowing controllers to continually focus their attention on monitoring the airspace instead of trying to comprehend the previous transmissions.

For H2, t-test results indicated that perceived task difficulty for the experimental group, mean rating 8.17 with a standard error of 2.065, was not significantly different than the control group, mean rating of 7.40 with a standard error of 1.802 ($t(68) = -1.665$, $p = .100$). The findings for H2 indicate that there is no significant difference in perceived difficulty of the task due to the use of the new technology. However, it should be recognized that the controllers were not required to conduct each dialog to its completion, only indicate through the mouse click when additional information was required.

For H3, t-test results indicated that recall in the experimental group, mean of 3.14 with a standard error of 1.52, was not significantly different than the control group, mean 3.20 with a standard error of 1.21($t(68) = 0.174$, $p = .862$). The findings for H3 indicate that there is no significant difference in memory recall due to the use of the new technology. The role of memory has been heavily researched in the air traffic environment and its importance concerning whether the controllers should retain memory or not has been debated. Some studies suggest that controllers deal with a constantly dynamic environment and therefore dynamic memory changes frequently. Hopkin (1980) stated that the ability to forget the information is as important as the ability to remember, allowing controllers to deal with just the necessary information required and avoid the workload associated with keeping memory. Even if doesn't improve memory recall, the text on screen might still be a good aid because controller can use the text to refresh his/her working memory.

For H4, t-test results indicated that distraction for the experimental group, mean 6.09 with a standard error of 3.381, was significantly higher than the control group, mean 4.57 with a standard error of 2.779 ($t(68) = -2.047$, $p < .05$). Therefore, a stepwise regression was run and the model indicated an adjusted R^2 of .093. This model included the scenario and additionally indicated that participants who self reported high ability to multitask and feeling comfortable with reading and listening at the same time experienced lower distraction. Due to the very low variance explained, definitive conclusions cannot be drawn for Hypothesis 4 at $F(69) = 4.537$, $p = .014$. Although distraction did appear to increase with the on-screen technology, no definitive conclusions could be drawn due to the very low variance explained in the model. This is an important finding since the low adjusted R^2 may indicate that while there was significantly more distraction in the experimental condition, the distraction may not have been due to the additional text. Other factors yet unknown may have also played a role.

Conclusions and Recommendations

There has been significant advancement in speech-to-text technology in recent years. The application of this technology is constantly increasing, specifically in military systems (Weinstein, 1991). The resulting systems provide the immediacy available from voice communications while providing persistence of information to aid deconfliction of simultaneously-broadcast communications and to serves to aid short-term memory.

This research represents the first step towards understanding the impact of speech-to-text technology for aiding the communication between air traffic controllers and the multitude of aircraft they control. In this study, we assumed that text transcription of pilots´ transmissions were completely accurate and measured the controller's need for repeat auditory communication, their recall of relevant information and their perceptions of the system, the difficulty of using the system. The results indicated that the controller's valued the system, would make fewer requests to the pilots for "say again" communications, and had similar recall. The controllers did not perceive a difference in difficulty with the reduction of text cueing, but were not required to complete their full cycle of communication tasks, associated with a "say again" communications. Therefore, it is possible that the controllers did not consider the time savings and reduction of tasks associated with the reduction in these tasks when rating difficulty. These results are encouraging and support the application of this technology to provide automated transcriptions of pilots' communication to enhance controllers' efficiency.

Despite the encouraging results, the participants' rated perceived distraction higher for the system involving text transcription than for their traditional system. This result provides a cautionary warning as it signals the possibility that the provision of the text transcription window could distract the air traffic controllers attention away from higher priority displays, such as the radar scope. As a result, additional research is recommended to understand the safe application of the proposed technology to enhance controllers´ efficiency. Although every effort was expended to recreate a realistic ATC workstation and environment, a more accurate interactive environment or training workstation could provide additional insight into this application. The use of an eye tracking to determine time spent reading the text, and therefore the time spent with eyes away from the radar scope is also recommended as it could provide insight into the impact that text cueing of pilot communications has on system safety.

Additionally, further research should investigate the effect of the reliability of text transcriptions on controller´s efficiency. Further, training procedures should be developed and included in future studies as the participants identified the lack of training with the technology as a barrier to the full utilization of the system.

This paper proposes an engineering tool designed to improve the cognitive process of the air traffic controller to support timely human decision making. An attempt was made to provide an overview of the cognitive factors that must be considered to provide a basis for the development of this tool to enhance efficiency and critical decision making. It is believed that by optimizing and providing all the information needed by the controller in a timely manner, the ATC system may benefit and that with highly optimized and correctly designed solutions, providing the potential to enhance the time-critical aspect of controllers decision making in an increasingly congested communication environment and potentially permit the controller to successfully control additional aircraft. In this specific application, controllers will be able to read what was said instead of just relying on auditory memory and comprehension. This advance could lead to increased efficiencies in the NAS throughout reduced frequency congestion and potentially error reduction associated with misunderstood transmissions.

This research was an exploratory study to determine the effectiveness and efficiency gaining potential of this new technology and whether further research is warranted. A number of contributions and implications of the use of the text technology were described. Applications in training and real-time air traffic are seen to be promising for the proposed technology. Additionally, future research is recommended to guarantee the advancement of the technology and its application to enhance the ATCS efficiency and safety.

Acknowledgements

The authors would like to thank the Air Force Air Traffic Control community for supporting this research. Without their support for the topic and their willingness to provide access to the participants at Kessler AFB, this research would not have been possible. Additionally, the authors would like to the 711th Human Performance Wing for their contributions of software and expertise. Finally, we would like to acknowledge the Federal Aviation Administration for providing the SIGNAL training software upon which the experiment was based.

References

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 85-104.

Cardosi, K. M., & Murphy, E. D. (1995). *Human factors checklist for the design and evaluation of air traffic control systems* (No. DOT-VNTSC-FAA-95-3.1). JOHN A VOLPE NATIONAL TRANSPORTATION SYSTEMS CENTER CAMBRIDGE MA.

Federal Aviation Administration, *FAA Aerospace Forecast Fiscal Years 2012 – 2032*, U.S. Department of Transportation, 2011

Finomore, V.J., Popik, D., Dallman, R., Stewart, J., Satterfield, K., & Castle, C. (2011) Demonstration of a Network-Centric Communication Management Suite: Multi-Modal Communication. Proceedings of the Human Factors and Ergonomics Society, 55, 1832-1835.

Hopkin, V.D. (1980). The measurement of the air traffic controller. *Human Factors*, 22, 547-60.

Hopkin, V. D. (1995). *Human factors in air traffic control*. CRC.

Parasuraman, R**.**, & Metzger, U. (2006). Effects of automated conflict cueing and traffic density on air traffic controller performance and visual attention in a datalink environment. *International Journal of Aviation Psychology*, *16*, 343-362.

Rantanen, E. M., McCarley, J. S., & Xu, X. (2004). Time delays in air traffic control communication loop: effect on controller performance and workload. *The International Journal of Aviation Psychology*, 14(4), 369-394.

Shorrock, S. T., & Kirwan, B. (2002). Development and application of a human error identification tool for air traffic control. *Applied Ergonomics*, 33(4), 319-336.

Stein, E. S., & Garland, D. (1993). *Air traffic controller working memory: considerations in air traffic control tactical operations* (No. DOT/FAA/CT-TN-93/37). FEDERAL AVIATION ADMINISTRATION, TECHNICAL CENTER, ATLANTIC CITY, NJ.

Tindall-Ford, S., Chandler, P., Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied*, Vol 3(4), 257-287. doi: 10.1037/1076-898X.3.4.257

Weinstein, C. J. (1991). Opportunities for advanced speech processing in military computer-based systems. *Proceedings of the IEEE*, 79(11), 1626-1641.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, 50:449 – 455.

Xing, J. and Manning, C. A. (2005). Complexity and automation displays of air traffic control: Literature review and analysis (No. DOT/FAA/AM-05/4). FEDERAL AVIATION ADMINISTRATION, OFFICE OF AEROSPACE MEDICINE, WASHINGTON, DC.

TRANSFERRING HUMAN FACTORS KNOWLEDGE FROM AVIATION TO DEVELOPMENT OF A WARNING SYSTEM FOR LANDSLIDE

Ioana Koglbauer, Brightline Avionics, Kalsdorf, Austria
Guenter Hohenberger, National Alarm- and Warning Center, Styria, Austria
Christian Steger, Institute of Technical Informatics, University of Technology, Graz, Austria

There is a successful history of transferring knowledge from aviation to other domains such as medicine (Thomas & Helmreich, 2002). In this study the ICAO multistage alerting service (2008) served as model for the specification of an early warning system for landslide. The early warning system is designed to monitor mass-movement data provided by GPS sensors, and to generate warnings and alarms to the National Alarm- and Warning Center of Styria, Austria. For the human factors specification of the new system a qualitative analysis was performed. Results are discussed with regard to applicability of human factors guidelines from aviation to development of systems for regional alarming centers which initiate and supervise disaster management activities.

The change in earth climate leads to an increasing number of torrential rain events and the consequence of mass-movement events with different dimensions. As Austria is located within the inner Alpine regions its infrastructure and population are exposed to a certain danger. By now there is no adequate early-warning-system for such mass-movement events available in Austria. As Lettieri, Masella and Radaelli (2009) remarked in a review of international disaster management studies, the use of technologies based on satellites, ground sensors and specific decision support systems have been too less investigated until now. In the Projekt GeoWSN a real-time monitoring system with "Wireless-Sensor-Network" with GPS positioning technique is being investigated.

The multistage alerting service for search and rescue described by ICAO (2008) and implemented by member states was used as model for the specification of three warning stages for landslide. Originally, ICAO defines the uncertainty (INCERFA), alert (ALERFA) and distress (DETRESFA) phases, and specifies activation criteria for each phase and corresponding action plans.

Furthermore, human factors issues for the human machine interface were identified based on aviation human factors design standards published by the Federal Aviation Administration (Ahlstrom & Longo, 2003) and reviews of disaster management literature (Collins & Kapuku, 2008; Lettieri et al., 2009; Simpson, 2008; Uniyal, 2008; Kapucu, Arslan & Demiroz, 2010).

Method

For the human factors specification of the new system a qualitative analysis was performed with participation of thirteen members of the development team, including authorized rescue agents, experts in geology, navigation, informatics, communication technology and psychosocial crisis intervention.

A questionnaire structure was elaborated based on following sources:

- Applicable disaster management regulations and guidelines,
- A review of functionality requirements for technical systems (Bayrak, 2009),

- Organizational issues involved in disaster monitoring and management (Kapucu, Arslan & Demiroz, 2010; Collins & Kapuku, 2008; Jaques, 2010),
- Usability of the system (Chin, Diehl & Norman, 1988),
- Alarm thresholds with respect to the balance between missed and false alarms (Parasuraman, Hancock & Olofinboba, 1997), and procedures for false alarms.

Results

Alerting Phases

The multistage alerting service described by ICAO (2008) was used as model for the specification of three warning stages for landslide. The warning system for landslide should be implemented in alpine areas where there is a risk of creeping slopes over a longer period of time. In this case expert observers cannot be assigned for a long term to monitor the area, and the experience showed that the assignment of non-expert observers often leads to false alarms and waste of valuable time and resources. Thus, the new GPS-based system shall be implemented to provide objective data which are processed by the system according to a set of area-specific thresholds which are set by geology experts.

Similarly to the ICAO (2008) alerting service, in our study three warning phases were defined, including triggers and action plans. Of course, our triggers and action plans were specific for landslide. In the first phase there is monitoring, but unlike in aviation, this phase can last for a long period of time if the acceleration of the mass-movement does not exceed a specified threshold. In the second warning phase the mass-movement reached 90% of the threshold value and expert observers are sent to observe the area. The concerned emergency organizations and population are informed to expect landslide and access in the danger area is restricted. In the third phase of alarming, there is an acute danger of landslide and the population and emergency organizations are alarmed to activate their previously specified action plans.

Interface Design

This section addresses the application of human factors system design standards from aviation published by FAA (Ahlstrom & Longo, 2003) to disaster monitoring and alarming systems.

A closer look at working conditions of the agents in the alarming center showed that, the same as pilots and air traffic controllers, they monitor a large amount of systems in a dynamic, distributed and real-time environment. However, their systems did not reach the level of standardization we face in the aerospace branch. Especially displays for warning systems are provided by a multitude of sources with various design philosophies (Figure 1).

The agents have short response times and are expected within minutes to identify dangers and warnings, and to deploy the respective action plans. It is obvious that the display design for disaster monitoring, same as aircraft cockpit design influences the performance of the operator (Trollip & Jensen, 1991). Thus, we found that early warning, anticipatory cues and multiple coding of display indications, and multimodality of warning and alarms were key human factors requirements.

The use of colors for coding the warning phases was similar to the aviation standard: red for alarm, yellow for caution, green when no danger is detected by the system. The freedom in

Europe with respect to color coding of disaster warning can be seen in some other systems which use purple for the highest alarm phase.

***Figure 1.* “Cockpit” of the authorized agents in the alarming center of Styria.** Various warning systems are displayed on different monitors.

***Figure 2.* A model of the graphic interface of the early-warning system for landslide.** The display shows multiple coding of the system status.

Another example illustrated in Figure 2 is the multiple coding of system status indications which use both colors (green and red) and symbols (e.g. x for missing data flags).

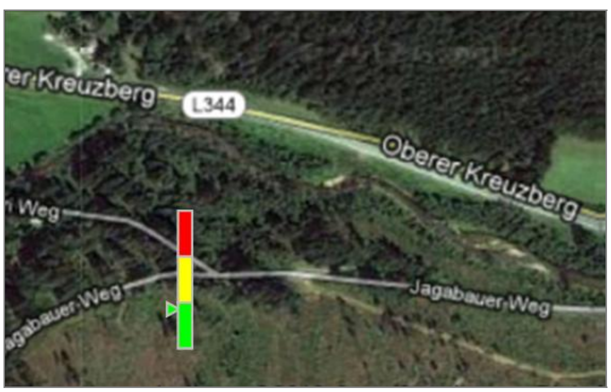


***Figure 3.* A model of the graphic interface of the early-warning system for landslide.** The display shows multiple coding of the warning phase.

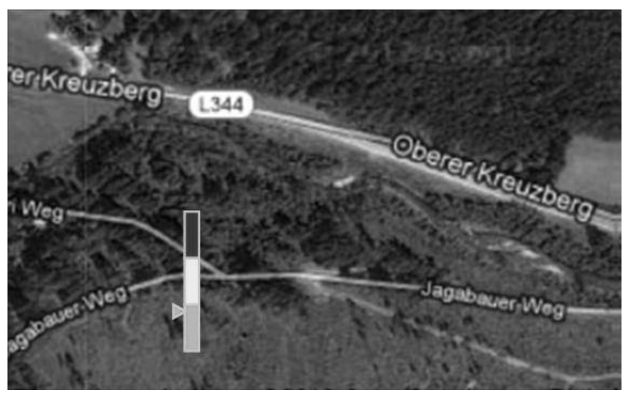


***Figure 4.* The same model of the graphic interface in gray.**

Furthermore, for displaying the warning phase we selected an indication similar to aircraft engine instruments (Figure 3), which indicates the actual warning phase and also allows anticipation. In this case the position of the arrow shows that the system currently does not detect any danger, but the arrow is near the threshold to the second warning phase indicating that a change might occur. Another benefit of this kind of indication is that it intuitively shows the degree of danger from low to high and it can be well interpreted independent of the color (Figure 4).

One reason for multiple coding is that although color coding has the benefit to speed up decoding of an indication (see Ahlstrom & Longo, 2003), there might be operators with color blindness who cannot rely on color codes alone. This issue is not so uncommon even among recreational pilots, but it is seldom considered by designers.

Furthermore, for warning and alarms a multimodal procedure was developed using besides visual display indications also aural cues, triggered by the communication equipment of the agents (e.g. SMS, Alarm-App).

Usability criteria of the human-machine-interface (HMI) of the warning system were specified in terms of functions, terminology and information, display modality, user concept and error tolerance which are not described here more detailed.

However, there were more human factors issues to be considered in developing the human machine interface of the system which required inputs from different experts who participated to the development: system functionality requirements and organizational issues. These are rather typical disaster management issues.

System functionality requirements

This section addresses system functionality requirements or what the agents should expect from the system in terms of accuracy, availability, redundancy, reliability, maintainability (Bayrak, 2009). For example accuracy of sensors was set at the range of 10 cm and data transmission failures of 5%. The capability of the system to identify and display data errors, as well as procedures for designing an error tolerant system, and abnormal procedures for technical failures that cannot be avoided were specified. Availability, derived from the time system is working divided by the time when the system exists between failures was set between 95% and 99%. Redundancy was specified in terms of redundant software and hardware to compensate especially for loss of components. Another critical feature to define was the system reliability, as lifetime of components or time to failure of a unit. Furthermore maintainability of the system was specified as a measure of the time within the system can be repaired after a failure.

Organizational requirements

This section addresses system features necessary for the organizations involved in warning and disaster management. From organizational point of view system performance factors were considered the connectivity to and the availability of a commonly accessible pool of information using a web interface (e.g. chat-room, messaging, photo-sharing). Furthermore, connectivity with another system providing demographical and strategic infrastructure information was considered.

Discussion

Generally results are supportive for the applicability of human factors system design standards from aviation to disaster monitoring and alarming systems. However, this should not be seen as an over-generalization. Disaster management is a growing research field with specific domains, actors and regulations. Besides the scientific approach, other factors and interest areas (e.g. political, economic, social) play a major role on interpretation, prevention and management of disasters.

References

Ahlstrom, V., & Longo, K. (2003). *Human Factors Design Standards (HFDS).* Federal Aviation Administration, Report No. DOT/FAA/CT-03/05 HF-STD-001.

Bayrak, T. (2009). Identifying requirements for a disaster-monitoring system. *Disaster Prevention and Management*, 18(2), 86-99.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings of SIGCHI* , 213-218.

Collins, M. L., & Kapucu, N. (2008). Early warning systems and disaster preparedness and response in local government. *Disaster Prevention and Management*, 17(5), 587-600.

ICAO International Civil Aviation Organization (2008). *International Aeronautical and Maritime Search and Rescue Manual (IAMSAR Manual).* Montreal, ICAO Doc 9731-AN/958.

Jaques, T. (2010). Embedding issue management as a strategic element of crisis prevention. *Disaster Prevention and Management*, 19(4), 469-482.

Kapucu, N., Arslan, T., & Demiroz, F. (2010). Collaborative management and national emergency management network. *Disaster Prevention and Management*, 19(4), 452-468.

Lettieri, E., Masella, C., & Radaelli, G. (2009). Disaster management: findings from a systematic review. *Disaster Prevention and Management*, 18(2), 117-136.

Parasuraman, R., Hancock, P. A., & Olofinboba, O. (1997). Alarm effectiveness in driver-centred collision-warning systems. *Ergonomics*, 40(3), 390-399.

Thomas, E.J., & Helmreich, R.L. (2002). Will airline safety models work in medicine? In M. M. Rosenthal, & K. M. Sutcliffe (Eds.). *Medical Error: What do we know? What do we do?* San Francisco, Jossey-Bass, 217-234.

Trollip, S. R., & Jensen, R. S. (1991). *Human factors for general aviation.* Jeppesen Sanderson Inc., Englewood.

Uniyal, A. (2008). Prognosis and mitigation strategy for major landslide-prone areas. *Disaster Prevention and Management*, 17(5), 622-644.

Acknowledgements

This study was performed as part of the research co-operation Project GeoWSN, founded by the Austrian Ministry of Transportation, Innovation and Technology (BMVIT) and Austrian Research Agency (FFG), KIRAS Number 832344.

We gratefully acknowledge Manfred Wieser, Katrin Landfahrer, Norbert Kuehtreiber, Daniel Koch, Roman Lesjak, Cornelia Forstner, Klaus Aichhorn, Philipp Berglez, Cornelia Aichhorn, Leander Hoermann, Marcus Dietl, Hannes Mock, Stephan Gether and Florian Schramm for their valuable inputs to this study.

APPLIED HUMAN FACTORS RESEARCH FOR THE TECHNICAL OPERATIONS ORGANIZATION OF THE FEDERAL AVIATION ADMINISTRATION

Edward Austrian
Fort Hill Group
Washington, DC USA

Edmundo Sierra, Jr.
Federal Aviation Administration
Washington, DC USA

The FAA is developing and implementing human factors standards within the Technical Operations domain of the Air Traffic Organization. TO personnel are responsible for installing, certifying, and maintaining NAS infrastructure and equipment. Application of standards within this domain is intended to improve human performance, contribute to the more efficient and effective maintenance of NAS systems, and facilitate the integration of human factors into operational systems. Over several decades, the FAA has independently and incrementally modernized TO systems leading to user-system interface diversity within and across systems and facilities. Based on human factors principles, TO requested that FAA human factors research develop human factors standards to create user-interface uniformity across TO systems. Based on a review of industry and academic literature from the aviation, nuclear, and communications domains, there is evidence that the application of standards may benefit training and human performance while potentially providing the Agency with noteworthy returns on investments.

The Federal Aviation Administration (FAA) Human Factors Division oversees the activities of the Air Traffic Control (ATC) / Technical Operations (TO) Human Factors Team. The Team is responsible for managing four research and development (R&D) portfolios that are driven by Destination 2025, the National Aviation Research Plan (NARP), the Air Traffic / Technical Operations Human Factors Strategic Research Plan, and the operational needs of internal FAA sponsoring organizations. Each R&D portfolio is composed of related requirements that aim towards improving a specific component of the National Airspace System (NAS). This paper details specific projects within the Advanced Technical Operations Systems (ATOS) R&D portfolio and the proposed benefits that may be achieved through the development and application of human factors standards within the TO domain of the Air Traffic Organization (ATO).

The Role of Technical Operations in the NAS and Shortfall Definition

The ATOS R&D portfolio aims to improve human factors contributions to the TO domain of the ATO. "The FAA TO Organization includes the centralized National Operations Control Center (NOCC), three regional Operations Control Centers (OCCs), Systems Operations Centers (SOCs) at Air Route Traffic Control Centers (ARTCCs) and large Terminal Radar Approach Controls (TRACON), and additional facilities at the local and regional level" (Chinoy

& Fischer, 2011). Within these facilities, TO personnel are responsible for the installation, certification, and maintenance of a wide variety of infrastructure, equipment, and systems. TO traditionally interacts with these systems through Graphical User Interfaces (GUI). When maintenance events are detected, coordination with, and prioritization of maintenance is determined by the local ATC facility prior to execution (Chinoy & Fischer, 2011).

Over several decades, the FAA has implemented independent, incremental improvements to TO systems. Documentation utilized by the FAA and system developers during those improvements addressed the incorporation of computer-human-interface design standards but did not bound a TO system developer to domain specific human factors standards—this resulted in complex and diverse graphical user interfaces within and across systems and facilities. "As a result, the likelihood of human error increased presenting the opportunity for unintended AT system outages and human performance inefficiencies" (FAA, 2012). In response to the identified shortfall, the FAA ATC/TO Human Factors Team and Technical Operations have partnered with industry to improve human factors requirements in system acquisitions. Among the products the partnership will develop and apply are the Graphical User Interface Standard, Graphical User Interface Style Guide, Technical Operations Maintenance Markings and Symbols Standard, and a Technical Operations Abbreviations Standard.

Due to the lack of uniformity and human performance data from legacy TO systems, this paper will propose human factors benefits and potential success criteria to be realized post-application of these human factors TO standards. The measures may be used to diagnose whether there is an opportunity to further improve performance, assess the effectiveness of the human factors solution, and determine whether there are opportunities for the community of practice to develop additional human factors interventions. Figure 1, below, provides a graphical overview of the proposed improvements, measures, and potential success.

The remainder of this document will further detail active research requirements for each of the aforementioned standards and corresponding literature review findings.

Figure 1. Human Performance Metrics

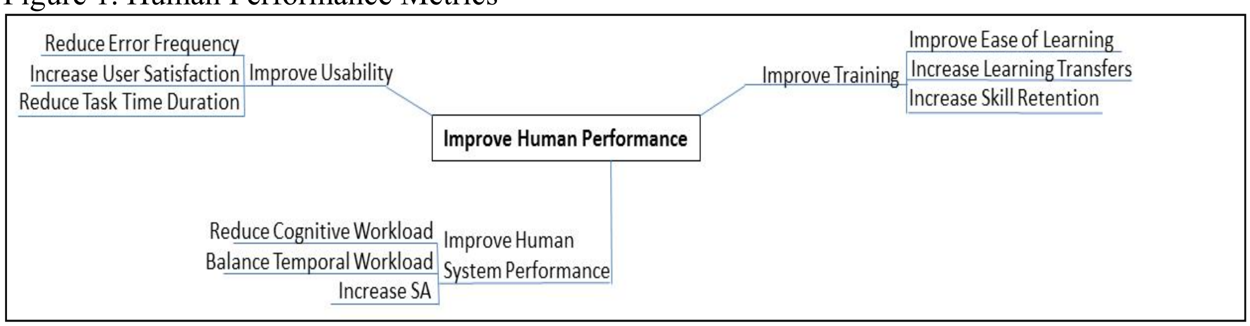


Figure 1 is a graphical overview of three main categories that support the improvement of Technical Operations human performance through the application of human factors standards. Each category (Improve Training, Improve Human System Performance, Improve Usability) are further defined by category specific measures. The measures associated with each category's success criteria are post product application goals that may be utilized to assess the need for future human factors interventions in TO. Figure 1 was adapted from the Task Performance and Human System Performance Metrics taxonomies (Pester-DeWan & Oonk, 2006).

Graphical User Interface Standard and Style Guide Requirement

Technical Operations is responsible for monitoring and controlling numerous NAS systems and equipment that directly influence air traffic system availability and NAS capacity through GUIs. TO has the ability to input information, receive information, and exert system control inputs through GUI interactions. These TO systems and their respective GUIs have been developed independently by different organizations. The independent and incremental development of these systems has created inconsistent user-system interfaces both within and across TO facilities. To prevent GUI inconsistencies during future system improvements or legacy technology replacement, the FAA has partnered with industry to develop a human factors TO GUI Standard.

Since it is critical that future system developers correctly apply the GUI Standard, the FAA has partnered with industry to develop a GUI Style Guide. The Style Guide shall link directly to the GUI Standard. The Style Guide is intended to eliminate any abstractions, ambiguity, or possible misapplication of the GUI Standard by developers. Application of both the Standard and Style Guide will ensure that future technologies in TO have a common look and feel to users.

Technical Operations Maintenance Markings and Symbol Standard

Symbols represent complex concepts in a succinct form that save space and are used to develop situational awareness in the performance of decision-making tasks. A user’s understanding of the meaning of something is closely connected with the task goals of the individual. From idea to implementation, it is important that developers incorporate human factors best practices in symbol design as a barrier to unintended human performance outcomes. Designs and design trade-offs developed in the course of acquisition of information and display systems for use in the NAS must include the user's information requirements and decision-making authorities and responsibilities. Systems must not be implemented in a manner that exceeds the user's cognitive capabilities and limitations in the context of the decision-making tasks (Narkevicius, 2012).

Symbols must be crosscutting to be effective. Advances in system designs may modify the role of maintainers and reinforce the need to convey crosscutting concepts and information. Different maintenance roles will need to communicate system status and availability information across levels of the organization succinctly, accurately, and at the granularity necessary for action at each of those levels (Narkevicius, 2012).

With the intent to further integrate human factors into operations, the ATC/TO Human Factors Team is partnering with industry to develop a Technical Operations Symbol Standard. The standard will address the creation, use, structure, and content of symbols, icons, markings, legends, text, and any other constructs conveying information on TO displays. The published standard will be applied as a requirements document for FAA TO system acquisitions.

As future systems evolve, there may be developmental or new concepts that are not covered by the Technical Operations Symbol Standard. Therefore, it is important for consistency that the FAA have evaluative guidelines for the development and approval of new symbols within TO. The ATC/TO Human Factors Team is partnering with industry to create evidence based evaluative symbol guidelines. The guidelines will contain: best practices for symbol design, defined criteria meriting the development of new symbols by vendors, as well as a formal vendor symbol application process and a formal approval process for TO.

Abbreviations Standard

Abbreviations are used on permanent markings, labels, and electronic displays on Technical Operations hardware and software. An abbreviation is a shortened version of a word or group of words formed by omitting one or more letters. In this paper, the authors use abbreviations as a collective term for acronyms, initialisms, contractions, and clippings. Regardless of the term used to identify the specific abbreviation, the shortened word or group of words is used on hardware and software to save space.

Lack of standardized Technical Operations abbreviations increases workload and likelihood of error. In medicine, some abbreviations are known to lead to misinterpretation and result in patient harm. To increase patient safety, there is an official "Do Not Use" list that applies to all orders and all medication-related documentation (Joint Commission, 2004).

The use of abbreviations on Technical Operations hardware and software is inconsistent. The primary reason for inconsistencies may be that a list of abbreviations to promote consistent use does not exist. For those abbreviations not appearing in the GPO Style Manual (2008), an engineer considering the use of an abbreviation must rely on his or her team's existing knowledge of over 3,000 abbreviations used in TO. Programs are unlikely to allocate a part of their very limited resources to reviewing existing TO systems for their use of abbreviations. Therefore, it is challenging to ensure that words have only one abbreviation, and abbreviations have only one definition.

Literature Review Findings

The development and application of human factors standards within Technical Operations is intended to act as a barrier for unintended designer errors leading to unintended operator outcomes (FAA, 2012). Human centric standardization across future TO system updates begins with providing industry user-friendly standards and guidance documents. Utilization and application of those documents by designers may improve end user human system performance, system usability, and training efficiencies. Additionally, there are potential program management and developer benefits to standardization—such as reduction in costs and the opportunity to reuse code (Nielsen, 1993).

User-oriented designs should allow expedient access to the status of individual components of a control system and their relationship with other components (Brookhaven National Laboratory, 2000). Implementation of soft controls is a technique that may be utilized by system developers to effectively utilize limited screen real-estate (Brookhaven National Laboratory, 2000). According to the 80-20 Rule, approximately 80% of users utilize a handful of an application's features, while only 20% use all or most of those features (Apple, 2012).

Consistent user-oriented designs, may improve a user's productivity resulting in higher throughput and a reduced number of errors due to system predictability. The smaller the number of errors and reduced learning times may also increase user satisfaction due to decreasing user frustrations (Nielsen, 1993). Interface consistency traditionally enhances a users' ability to effectively transfer user skills from one system to another, leading to ease of learning and use—thus potentially lowering training costs (Nielsen, 1993). Corroborating this statement, Polson

1988, “In several studies, consistency reduced training time to between 25-50% of that needed for inconsistent interfaces.” Application of user-centric designs will promote the ability for users to learn programs faster due to elements looking and behaving the same (Apple, 2012).

Conclusion

Application of standards within Technical Operations is intended to improve human performance, contribute to the more efficient and effective maintenance of NAS systems, and facilitate the integration of human factors into operational systems. Human factors requirements are intended to ensure that NAS equipment utilized by FAA personnel are easy to operate, maintain, and train (FAA, 2011). The aforementioned standards and proposed success criteria may be used to diagnose whether there is an opportunity to further improve performance, assess the effectiveness of human factors solutions, and determine whether there are opportunities for the community of practice to develop further interventions.

Acknowledgements

The authors would like to acknowledge the FAA’s Human Factors Division (ANG-C1) for funding this effort as well as the Technical Operations research performers.

Disclaimer

The opinions expressed are those of the authors and do not represent the Federal Aviation Administration (FAA).

References

Apple. (2012). *OS X Human Interface Guidelines.* Retrieved from http://developer.apple.com/library/mac/#documentation/userexperience/conceptual/applehiguidelines/Intro/Intro.html#//apple_ref/doc/uid/TP30000894-TP6.

Brookhaven National Laboratory. (2000). *Soft Controls: Technical Basis and Human Factors Review Guidance* (NUREG/CR-6635). Washington, DC: Nuclear Regulatory Commission.

Chinoy, S. & Fischer, D. (2011). Advancing Situational Awareness for Technical Operations. Proceedings from the *56th Air Traffic Control Association (ATCA) Annual Conference*, Maryland, USA.

FAA. (2011). *Guidelines for Human Factors Requirements Development*. Retrieved from https://www.hf.faa.gov/hfportalnew/SAE.aspx

FAA. (2012). *Air Traffic Control / Technical Operations Human Factors Strategic Research Plan.* Retrieved from https://www.hf.faa.gov/hfportalnew/Uploads/gcreighton/ATC%20TO%20HF%20Strategic%20Plan%20November%202012%20Version%201.0.pdf

Government Printing Office. (2008). *Style Manual: An Official Guide to the Form and Style of Federal Government Printing*. Retrieved from http://www.gpo.gov/fdsys/pkg/GPO-STYLEMANUAL-2008/pdf/GPO-STYLEMANUAL-2008.pdf

Pester-Dewan, J., Oonk, H. (2006). Human Performance Benefits of Standard Measures and Metrics for Network-Centric Warfare. Proceedings from the *2006 Human Factors Issues in Network-Centric Warfare*, Sydney, Australia.

Polson, P.G. (1988). The Consequences of Consistent and Inconsistent User Interfaces. In Guindon, R. (Ed.), *Cognitive Science and its Applications for Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Joint Commission. (2004). *Facts About the Official "Do Not Use" List*. Retrieved from http://www.jointcommission.org/assets/1/18/Do_Not_Use_List.pdf

Narkevicius, J. (2012). *Symbols and Markings for Technical Operations: Glossary of Symbols*. Unpublished manuscript.

Nielsen, J. (1993). *Usability Engineering*. San Francisco, CA: Morgan Kaufmann.

GIVING A FACE TO AIRLINE CUSTOMER SATISFACTION: A GRAPHIC APPROACH

Clay Wildt
Dr. Erin E. Bowen
Dr. Brent D. Bowen
Purdue University
West Lafayette, Indiana

Historically, research ranking the major commercial air carriers in the U.S. has been based on subjective perceptions, satisfaction, and attitudes. Building upon 21 years of work with the Airline Quality Rating (AQR), the present study moves beyond basic descriptive information of air travelers to identify patterns and relationships in the way consumers view this technologically advanced environment. Development of such a model allows key players in the industry to improve their understanding of the prime drivers and perceptions of passenger behavior. Implementation of a subjective element, the Wong-Baker Faces Pain Scale, will allow frequent fliers the ability to codify their feelings and emotions towards airline flying experiences. A crucial connection will be made between subjective perceptions measured through survey responses, and the formula-driven Airline Quality Rating; a graphic reference of each airline's perceived quality will be offered in the form of an emotional face.

The purpose of this work is to integrate quantitative ratings of airline quality with qualitative survey results to produce a graphic approach appropriate for public dissemination. No longer will the traveling public have to sift through hundreds of individual airline reviews, commentary about bad customer service experience or try to compare airlines' scores in an already difficult to understand realm of side-by-side comparison.

Literature Review

Using the Airline Quality Rating (Bowen & Headley, 2012) as a baseline to construct a new model for airline consumer satisfaction is a solid foundation, but consumer opinion is not completely reflected in the numbers and statistics reported by Department of Transportation (DOT) data (Rhoades, Waguespack, & Treudt, 1998). Extending research that has been conducted in other service industries (namely banking, pest control, dry cleaning, and fast food), this new model will venture into the realm of U.S. domestic airlines. Exploring attitude-based conceptualizations as opposed to traditional surveys, which are weighted according to researchers' interests or simply weighted equally, will produce a model that is more closely aligned with consumers' expectations (Cronin & Taylor, 1992).

Over 20 years ago, scholars and researchers indicated that service quality and satisfaction are mutually exclusive in the eyes of the customer (Cronin & Taylor, 1992); this revised model resists traditional logic by asserting the "oneness" of the two concepts, especially from the perspective of airline travelers. Service quality, as defined by each individual passenger, can be equated to satisfaction; if the airline meets or exceeds quality standards set by the passenger (internally), the customer will feel satisfied, and possibly delighted.

In designing a completely new model for gauging airline customers' perceptions of quality within the industry, Reeves and Bednar (1994) set the cornerstone by expressing their opinion, "quality is whatever the customers say it is" (p. 427). Expanding on this concept, they further assert that quality moves past a philosophical argument to a practical one with implications in every industry. Conforming to the laws of supply and demand, only the customer (who is a resident and user in the marketplace) can articulate the ultimate quality of a service as it meets their immediate or anticipated need (Reeves & Bednar, 1994).

Other scholars in the area of service quality recommended that a weighted model of customer satisfaction be created. For example, one study proposed a model that asked a pair of questions for each aspect of quality; the first hinged on the individual's perception that a company *should* have or provide a specific product or service, followed immediately by a question asking if said company actually *has* the aforementioned product or service (Lewis & Mitchell, 1990). This weighted model touches on individual definitions of quality and service, and rates companies according to personal perception. The authors further assert, "if a graphic scale were to be used, it would give additional validity to the use of parametric statistics" (Lewis & Mitchell, 1990, p. 15). The new model proposed here will combine elements first recommended above; having a ranking for relative importance followed closely by the actual service rating customizes each score to precisely fit the needs of each customer. Not only are numerical values obtained from the research, a graphic depiction of satisfaction is also offered in a scale of emotional faces.

Defining Quality

Scholars and researchers have had a difficult history in defining quality; many different definitions exist, all of which could be appropriate given various situations. Some commonly accepted definitions as cited in Reeves and Bednar (1994) include:

- value (Abbott, 1955; Feigenbaum, 1951)
- conformance to specifications (Gilmore, 1974; Levitt, 1972)
- conformance to requirements (Crosby, 1979)
- fitness for use (Juran, 1974, 1988)
- loss avoidance (Ross, 1989)
- meeting and/or exceeding customers' expectations (Gronroos, 1983; Parasuraman, Zeithaml, & Berry, 1985)

Because the airline industry resides in the service sector as opposed to the manufacturing sector, quality cannot be measured in terms of bad parts per thousand, number of returned products, or even warranty claims. Instead, service firms, airlines included, should define quality through the eyes of their respective customers. Reviewing the list above, it only seems appropriate to assign "meeting and/or exceeding customers' expectations" as the defining metric for airline quality.

Each individual passenger has a unique perspective on what quality means within an airline. For example, some travelers are more concerned with getting to their destination with the lowest fares and least amount of fees. Others expect extravagant meal service and include the flight in the overall travel experience. Still others simply want to be treated fairly and have a comfortable experience during their flight. This very personal definition of quality is unique to each traveler; finding an effective way to depict individual quality definition is key to constructing a valid model of airline quality.

Existing Rating Systems

There are numerous academic and business based models currently available to rate and rank the quality of airlines both domestically and abroad. Each model takes a different approach to measuring quality and uses a variety of metrics to capture the feelings of passengers, both quantitatively (such as rankings, performance numbers, etc.) and qualitatively (through survey results, review forums, etc.). Listed below are some of the most widely distributed and relied upon systems to gather and disseminate information about quality within the airline industry.

Airline Quality Rating (AQR)

First published in 1991, the Airline Quality Rating was conceived by Drs. Bowen and Headley at Wichita State University. The report has been published annually for the past 21 years and has drawn significant media attention. Using a weighted-average formula, the model draws on measures taken from the U.S. Department of Transportation *Air Travel Consumer Report.* The following metrics are used in the master formula: on-time performance (OT), denied boardings (DB), mishandled baggage (MB), and customer complaints (CC)—this category contains items such as flight problems, oversales, reservations, ticketing, boarding, fares, refunds, baggage, customer service, disability, advertising, discrimination, animals, and other complaints (Bowen & Headley, 2012).

A more recent component to the Airline Quality Rating is the addition of the Airline Passenger Survey (APS), which captures qualitative and quantitative data in the form of open-ended inquiries and Likert-style questions where passengers can relay their positive and negative experiences to researchers. While the AQR and APS seek to convey passengers' feelings and attitudes toward airline travel, a graphically-based report to convey the data to consumer has yet to be produced (Bowen, Bowen, & Headley, 2012).

Zagat Airline Survey

Another long standing indicator of airline performance is the Zagat Airline Survey. Started in 1990, the survey collects data from more than 8,000 frequent fliers annually. The main indicators of performance in the Zagat survey include comfort, service, food, and website efficiency (ease of use, booking system, etc.). It should be noted that the following categories are also taken into consideration and rankings formed: value, timeliness, check-in, luggage policy, and in-flight entertainment. Each of the main indicators are rated along a 30-point scale, with airlines being rated in premium and economy areas. One of the most interesting deliverables for the Zagat Airline Survey includes the creation of an "array of tables" that outlines the demographics and preferences of the frequent fliers surveyed. In addition, quality ratings for U.S. airports are identified; while this may be outside the scope of determining airline quality, it should be noted that the airport environment has a measureable impact on passengers before they ever get to the passenger/airline interface (Zagat Survey, LLC, 2010).

SkyTrax Airline Review and Rating

A more informal system to measure airline quality exists at www.airlinequality.com (through SkyTrax). This system has two components; the first is constructed much like a message board or TripAdvisor© review interface. Users can enter a numerical rating from 0-10 to rate their experience, while also weighing-in individually on the following areas: value for money, seat comfort, staff service, and catering. A final field exists to indicate whether the passenger would recommend the airline to others (which can be ticked yes or no). Even though these reviews are not currently being used for any qualitative data analysis, they would serve as a wealth of information to improve quality for individual airlines, or even specific routes within an airline's structure.

The second component to the SkyTrax website is an airline rating section, which rates all carriers on a scale of 1 through 5; each airline is then assigned a "star value" representative of their respective survey scores (Plaisted, 2012). Individual carriers can also become a "Quality Approved Airline", which involves a rigorous audit that encompasses more than 750 unique areas of product and service quality. This standardized audit was developed more than 20 years ago and still stands as a global benchmark for quality in the airline industry (Plaisted, 2012).

Method

Data Source

In creating a new model, frequent fliers who provided their email address while completing the AQR were polled; these individuals were already familiar with the goals and style of the survey. This also creates continuity between the AQR, Airline Passenger Survey (APS), and the new model, as the opinions expressed by the subjects should be somewhat similar (since they have provided responses for the APS analysis).

The new aspect of the model focuses on adding a component to the existing AQR questionnaire. The questions use a style similar to the Customer Perceived Value (CPV) scale that is widely used to ascertain loyalty to specific companies, brands, or products (Evans & Lindsay, 2008). This model generates a quality score that will differ for each individual traveler, based on which attributes they find to be most important when traveling by air. After the individual ranked each item of importance, he or she rated the quality of each item on a 5-point scale. The advantage to using this type of system over a typical Likert Scale is the personalized nature of each individual review. Rolling these quantitative scores into graphic face indicators, travelers will have both quantitative ranking scores and a set of graphic indicators for comparison.

Data Examination

After the 11 day survey window elapsed, responses were compiled and analyzed. The first step is to build a discrete score for each respective airline. From there, researchers assigned each airline their own emotion face, adapted from the Wong-Baker Pain Scale (Wong-Baker Faces Foundation, 1983). A brief outline of the process follows, with attention to the calculations necessary in each area.

To get a final score, the relative importance (which will be a ranking, 1-5; 5 being the most important) was multiplied by the quality score (1-5; 5 being the highest quality) to produce a unique score for each individual passenger. This score was divided by the total possible points (seventy-five) to yield a percentage. Finally, the percentage was converted to a raw score out of a possible five points. When the final score was calculated, the value was added to the respective airline's collection of ratings; a mean score for each airline was calculated and an emotional face assigned.

Procedure

After initial data collection, raw data from Qualtrics was exported into Microsoft Excel, and then sorted into appropriate columns. A mean score for each airline was calculated, rounded to the nearest whole number, and then assigned an appropriate emotion face.

After the data was analyzed, there were two final products that could be used by researchers, scholars, and industry leaders. First, each airline has a score that corresponds to the average passenger ranking and rating scale previously introduced. Second, an emotional face (Adapted from the Wong-Baker Pain Scale) was attached to each airline signifying their "feel" from customers. In all, this graphically-based model will provide a "dashboard" of sorts for passengers to compare airlines.

Results

Survey results were collected for a total of 11 days; during this time, 334 responses were recorded (from about 7000 solicited email addresses). Of the respondents, 82% identified themselves as males, and 60% were reportedly between the ages of 42 and 65. When asked about their most recent airline experience, 21% of frequent fliers had flown with Delta Air Lines, 19% with United Airlines, 16% with Southwest Airlines, 13% with American Airlines, and 32% reported flying with other airlines. It is important to note that Mesa Airlines, Atlantic Southeast Airlines, and SkyWest Airlines did not have any respondents.

When asked to rank which items of the airline travel experience were most important, the answer chosen most frequently was fare prices/fees. On the other hand, a majority of respondents indicated that baggage handling (such as lost or damaged baggage, carry-on limitations, etc.) was the least important aspect of air travel. Rating quality of services yielded interesting results; customers found the highest quality in airlines' customer service (including ticket counter employees, gate agents, flight and cabin crew, general hospitality, etc.). The lowest rated aspects were airplane comfort (including in-flight entertainment, food and beverage service, and seat comfort).

Thorough analysis of the data yields a list of airlines, ranked by score. Table 1 (below) depicts two important lineups. First, the table on the left indicates each airline that received at least one survey response, as well as its respective score. The rightmost column depicts how many respondents identified each particular airline as their most recent carrier. Since many carriers had very few responses, a certain amount of bias is introduced into the model. For this reason, a separate table is shown, filtering out the airlines that received only a few responses. Only those carriers who received at least 10 responses are shown in the rightmost table.

Table 1.
Airline scores (left), filtered to include n≥10 (right)

Airline	Score	n
Hawaiian Airlines	4.24	3
Frontier Airlines	4.02	3
JetBlue Airways	4.00	9
Alaska Airlines	3.95	17
Southwest Airlines	3.88	52
American Eagle	3.83	2
Air Tran Airways	3.80	10
Delta Air Lines	3.48	71
United Airlines	3.39	65
Continental Airlines	3.39	6
American Airlines	3.25	43
US Airways	3.23	21
Mesa Airlines	2.27	1

→

Airline	Score	n
Alaska Airlines	3.95	17
Southwest Airlines	3.88	52
Air Tran Airways	3.80	10
Delta Air Lines	3.48	71
United Airlines	3.39	65
American Airlines	3.25	43
US Airways	3.23	21

Respondents who rated their customer service experience as a 5 usually identified with Southwest Airlines (22%). Interestingly, Southwest also captured 34% of those who rated fare prices and fees as a 5. Operating under a customer-centric, low-cost structure, the airline continues to attract frequent fliers who enjoy being treated well without the exorbitant ticket costs.

Discussion

Before the results of this study are thoroughly dissected, it is important to note the limitations of the project at hand. Since a convenience sample was used to collect data, it is much more likely for respondents (frequent fliers) to identify with larger carriers. Smaller, regional airlines generally have fewer frequent fliers than legacy carriers. Also, since participants chose respective airline, the distribution of responses was not evenly distributed. For example, 3 airlines did not have any responses, while another 6 carriers had less than 10 submissions. With so few responses, the data could easily be skewed by disgruntled or overly delighted passengers.

For the aforementioned reasons, analysis will focus on the list of carriers with 10 or more survey responses. Alaska Airlines, Southwest Airlines, and Air Tran Airways captured the top 3 positions with scores consistently at or above 3.80. These respective airlines also placed well in the Airline Quality Rating (AQR) and Airline Passenger Survey (APS). For example, Alaska Airlines has been in the top 5 positions in the Airline Quality Rating (AQR) for the past 2 years, and is often touted as a customer-friendly alternative to the more mature legacy carriers (Bowen & Headley, 2012). While Southwest rates lower in the AQR, it consistently captures the title of 'preferred airline' and most 'passenger-friendly airline' in the Airline Passenger Survey (APS) (Bowen, Bowen, & Headley, 2012). Finally, Air Tran maintains a lead in the AQR, not falling below 3rd place during the past 7 years. The findings of this survey are consistent with passengers' perceptions of airline quality as captured by the AQR and APS (Bowen & Headley, 2012).

Respondents were also given the opportunity to provide additional comments about their experience; many travelers took this opportunity to express their extreme discontent for how they were treated, and often at the industry at large. Reponses cover a wide range of topics including seat comfort, customer service "horror" stories, fare/fee complaints, as well as a variety of other comments. The overall tone of responses seems to point to an unfortunate lack of feedback mechanisms within the airlines. Some customers even assert that their concerns were not adequately addressed when the issue was brought to the attention of managers or customer service representatives.

The flagship output of this analysis is a graphic depiction of each airline's score; over time, travelers can refer to a range of scales to see if a particular airline's service quality has improved or declined. For this particular sample, each airline was assigned a score between 2 and 4 (no airlines qualified for a 1 or 5). Three select airlines and their respective facial representation appear below (Figure 1).

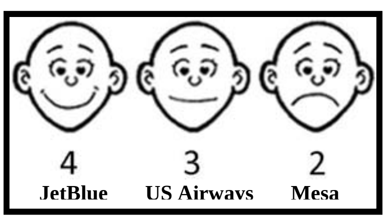


Figure 1. Graphic depiction of three select airlines

Using a larger, more comprehensive sample of frequent fliers, a list of all domestic airlines could be constructed. General passengers would be able to use this list as a "quick reference" to determine how other passengers, particularly frequent fliers, felt about travel on each respective airline. Even though the overall ranking scheme is similar to others that are current being used across the industry, the introduction of a graphic component will aid travelers by offering them a simpler, less 'number intensive' scale to compare their travel options.

Finally, the use of this model across time could lead to numerous, more intense, studies of passenger satisfaction during different events such as airline mergers, economic recessions, and bankruptcy restructuring. Researchers could easily track perceptions before, during, and after these events to report how customers perceived the airline differently, and how customer service was affected. Extending the application to airline management, stakeholders within the company could use the information to predict how customers will respond to certain events, and plan accordingly. For example, during a merger, management could increase on-board amenities and ramp up attention to customer service to offset the confusion and uncertainty associated with the merger. Knowing how customer satisfaction levels will fluctuate with shocks to the airline industry is a powerful tool in the arsenal of decision-makers; proactive steps to reduce a sharp slide in passengers' perceptions of the airline could help to counter falling revenues.

Conclusions

Using a variety of existing airline quality surveys combined with innovative components, this new model of measuring customers' satisfaction with their respective airline experiences provides an output that has never before been explored in a similar model. The ranking and rating system allows researchers to pinpoint individual components of quality as perceived by the consumers themselves. A traditional table of rankings combined with graphic depiction of each airline's quality will be presented to travelers so that they can quickly glance through the various domestic carriers and make travel plans accordingly.

References

Bowen, B. D., & Headley, D. E. (2012). *Airline Quality Rating.* Retrieved from http://www.airlinequalityrating.com/reports/2012aqr.pdf

Bowen, E. E., Bowen, B. D., & Headley, D. E. (2012). Frequent flier perceptions & air travel satisfaction: The airline passenger survey 2012. Unpublished manuscript, Department of Aviation Technology, Purdue University, West Lafayette, Indiana.

Cronin, J. J., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing, 56*(3), 55-68.

Evans, J. J., & Lindsay, W. M. (2008). *Managing for quality and performance excellence.* Mason, OH: South-Western Cengage Learning.

Herklots, T., & Sampogna, N. (2010). *Zagat surveys ninety major world airlines* [Press Release]. New York: Zagat Survey, LLC.

Lewis, B. R., & Mitchell, V. W. (1990). Defining and measuring the quality of customer service. *Marketing Intelligence & Planning, 8*(6), 11-17.

Plaisted, E. (2012). The SKYTRAX world airline audit (SAQA). Retrieved from http://www.skytraxresearch.com/General/audit.htm.

Plaisted, E. (2012). The world airline star ranking programme. Retrieved from http://www.skytraxresearch.com/General/ranking.htm

Reeves, C. A., & Bednar, D. A. (1994). Defining quality: Alternatives and implications. *Academy of Management Review, 19*(3), 419-445.

Roades, D. L., Waguespack, B., & Treudt, E. (1998) Service quality in the US airline industry: Progress and problems. *Managing Service Quality, 8*(5), 306-311.

Wong-Baker Faces Foundation. (1983). Wong-Baker FACES Pain Rating Scale. Retrieved from http://www.wongbakerfaces.org/

Zagat Survey, LLC. (2010). *Zagat Survey Summary: 2010 U.S. & international airlines survey.* Retrieved from http://www.zagat.com/sites/default/files/20101129_airlines.pdf.

The impact of cognitive psychology in minimizing human errors

Abstract

Working towards zero accidents – Experience from education and supervision of pilots and air traffic controllers (ATC) from the Royal Danish Air Force gives us a model on how to apply the latest clinical psychology methods and research, and combine it with Human Factor models.

What is behind accidents? Errors, so can you talk about zero errors? Everyone knows that this is utopia; the question is rather how we can understand errors, minimize errors and minimize the effect of errors. The goal of this paper is to contribute to a theoretical and practical understanding of how to use the research results of clinical psychology methods, hence the elements that you can transfer to a teaching situation to provide tools to be able to handle errors. In light of research findings from the clinical cognitive behavioral psychology, it is pertinent to examine the transfer values for teaching pilots and ATC. The following methods, concepts and materials have been used:

• Cognitive model of the mind
• Acceptance
• Focusing mentally and Visualization
• Reformulation

The clinical research findings presented in the paper is based on a literature review. Results in relation to transfer values presented in this paper are based on written and oral evaluation from the pilots and ATC after teaching. Therefore, the scaling results cannot be understood as evidence-based, instead, it is seen as an indication of how strategies from clinical psychology can be used in education of pilots/ATC to influence the self-perception and hereby reduce the amount and the impact of errors for the participants. Therefore, the paper can be used as an inspiration to practical use and further research.

Introduction

Before I started working in the area of Aviation Psychology I used several different treatment methods working as a clinical psychologist. Therefore, I have been very interested in the evidence of treatment methods. The idea to look into a transformation of methods from the world of clinical psychology treatment to the world of Aviation Psychology, emerged when I was teaching a course in psychology to pilot trainees within the Danish Defense. During the course, it became clear that some of the students feared making mistakes during flights, and they wanted to discuss how they could prevent errors. My interest to see how strategies from clinical treatment methods could be converted into teaching situations with the focus of how to handle mistakes and errors grew from these discussions. Now, the teaching has developed to not only relate to pilot students but also to students within Air Traffic Controlling (ATC), pilots and ATC's in various courses of continuing training (including in their CRM/TRM, instructor- and aviation safety training courses).

The reason I identified the relationship between clinical treatment and minimizing human errors was the basic element of psychology in both subjects. An essential part of clinical therapy is giving the client an understanding of the perception he has of himself as well as of the surrounding world, and furthermore making the client see how he can think or understand himself in different ways. The underlying idea of this education is to provide the client the strategies to take action, obtain a better capability to understand himself and his surroundings, and therefore also the possibility to change his reactions, e.g. in situations that may occur during a flight. It is especially this connection between treatment and education that makes it extremely interesting and relevant to develop and explore the effects of using methods from clinical therapy in Aviation Psychology.

Evidence-based treatment

There is a long tradition within therapy to study the evidence of the treatments. In 1999 Hubble, et al. published a meta-analysis based on the last 40 years of research and they summarized what worked in therapy. The results identified some general factors or basic elements which should be in place in the treatment to obtain a positive result, meaning the client would get better after therapy. Hubble et al. concluded that four general elements were present, (1) Extratherapeutic Factors (the client factor) which accounted for 40% (2) the Relationship Factor that accounted for 30%, (3) the Hope and Expectations Factor which accounted for 15% and (4) the Model and Technology Factor that accounted for 15%.

Extratherapeutic Factors include the client's personality, intellectual level, motivation, mental strengths and weaknesses, values, resources, potential, experience, etc. The second largest factor, the Relational Factor, is based on the client's perception of the therapist and includes understanding, acceptance, warmth, authenticity, etc. The Hope and Expectations Factor represent the client's expectations and the hopes of the possibility of development and improvement of the therapy. The last factor, the Model and Technology Factor (described below), includes the theoretical background a therapist is working with in therapy.

The focus is on the client's resources and how the person wants to experience the therapy. Does this mean that the therapist cannot affect the process? No. As described by Morawetz in *"What works in therapy? What Australian Clients Say"* the therapist has many opportunities to improve and strengthen client opinion through various strategies. This can be done in different ways, for example (1) by the therapist assessing the client's strengths and resources (2) asking the client to describe the problem and ask what solutions the client sees (3) by having the therapist focus on present and future solutions instead of past problems the client has had, or (4) by choosing a treatment appropriate for the client and (5) the therapist being is genuine and accepting during treatment. Therefore, all in all, one can say that the most important thing is for the therapist to meet the client where the client is.

When examining the Model and Technology Factor, what kind of results do we see? Cognitive Behavioral Therapy (CBT) is one of the most studied and evaluated approaches within therapy. Butler et al. states in their article, that between 1986 and 1993 120 studies had been conducted and now 325 studies on the effect of CBT effect have been published. One possible explanation for the high number of studies stated by Butler et al. is, relates to the positive results of this method of therapy regarding depression (this form of therapy started out for people suffering a depression) and hence researchers wants to see if CBT can be transferred to other areas. Later research has shown that CBT therapy also is highly effective in treating general anxiety, panic, social phobia and PTSD, and positive tendencies have been identified in many other areas (p. 17-31).

The relationship between evidence-based treatment and education#

What kind of parallels can be drawn when transferring evidence-based treatment into education for pilots and ATC? King describes in the chapter "Teaching" in the book "Aerospace Clinical Psychology" dimensions a teacher should considerate and should draw attention to when teaching pilots. The following are some examples from King's book regarding these elements: *"Above all else: know your audience! Prepare so that you appear spontaneous, know what you're talking about, use aviation metaphors, be funny"* (p.47-52). The parallels to Morawetz strategies of how to improve the relationship between client and therapist are clear, thereby, what is important for a positive outcome in treatment is therefore also important in teaching.

Of course, there are differences between teaching and therapy, but I believe that in both areas we see some basic elements or general factors that whether you are a therapist or a teacher are relevant to getting your message through to your audience. Hence, one must pay attention to those general factors and what tools you posses as a teacher to get your students attention. Given we have the students' attention and interest in place, we know that a large part of the teaching of pilots and ATC within Human Factors and CRM/TRM, is partly based on experiences, events, theories of human-machine interfaces, human-human interaction, basic theories within psychology and so on. Therefore, it is also important to look at how the understanding within psychology has developed during the last few years and hence also the forms of therapy, and how and what will be meaningful to transform to our target audience.

Free describes in his book *"Cognitive Therapy in Groups"* that he consider the CBT Group Therapy as a “psycho-education-group” or an "evening class on cookery” in which the focus is on teaching the clients new strategies for action (p. 40-41). How Free defines the CBT group, is of course also what you want in teaching, to provide students with knowledge and making them able to use this knowledge. The cognitive approach aims to give the client the tools to become their own "therapist", which is relates well to the aim of teaching. The connection between CBT and efforts to minimize human errors is that by becoming your own "therapist" or “teacher”, you will acquire both a new understanding as well as the tools to be able to react differently to your perception of a particular situation and hence avoid, minimize or stop a sequence of errors.

Cognitive therapy's theory

Cognitive therapy was developed by Aaron T. Beck. The basic understanding of the cognitive therapy is that the client in collaboration with the therapist explores the client's perception of herself and the outside world. Symptoms of the client's condition can be expressed through cognitive, emotional, physiological or behavioral responses. The symptoms reflects the perception the client has of herself, especially underlying assumptions expressed by the client that reflect automatically activated negative thoughts, often without the client being aware of this. The negative automatic thoughts, based in the client's underlying assumptions about herself, results in the individual's personal schedules, ie. her own understanding of herself. The cognitive theory suggests that the individual has some early learning and experiences that have led to the development of some dysfunctional schemes of the self and the surrounding world, but there are often some critical events that trigger development a mental illness.

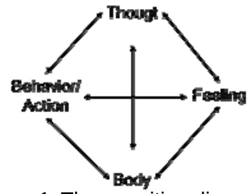


Figure 1. The cognitive diamond.

One cognitive model, by some named the Cognitive Diamond, looks at the individual's thoughts, emotions, body, behavior and the relationship between these elements. The thoughts a person can have in a given situation can lead to a feeling and/or a physical reaction which can result in more negative automatic thoughts and so on. In other words it becomes self-reinforcing. An example: During a debriefing, a student pilot is informed by his instructor, that he made an error during a flight. Depending on the student's self-perception and his past experiences, this situation can evolve in many different ways. If the student does not understand what he did incorrect, but does not dare to ask (passive behavior), maybe because the student thinks "*I am also too stupid to understand*" (thought) which results in uncertainty and anxiety (feeling). When a person experiences this, the body reacts with heart banking and sweating (body), which in turn leads to the idea that he will never become a pilot (thought). The learning attained by the student is about the student's personal schedules confirming that he is incompetent. The result is that the education regarding the error fails, and as a side effect, but an important one, the instructor beliefs that the student has understood the education and embraced it. When the student makes the error again the instructor becomes irritated (feeling) and thinks *"this student cannot learn this, it is going too slow, he will never become a pilot"* (thought). The instructor begins to look for errors regarding the student (perception) which in turn confirms the instructor's opinion, and ultimately it becomes self-validating for both the student and the instructor.

The cognitive model's understanding of thoughts includes the individual's values, rules of life, opinions, beliefs and motives. Regarding feelings, the model includes the six basic states of the human mood: happiness, sadness, anger, anxiety, astonishment and disgust. Regarding body, the model includes the energy level and the stress level, and regarding the last of the four dimensions, the behavior, the model includes skills (what you can do) and the habit (what you actually do). When dealing with the situation, the thoughts, the feelings, the body reactions, the behavior and the consequences of it all, you are looking into the past and present. To find solutions, we must also look at the need for change, ie. look at the person's goals and resources to achieve the desired result. Hence, you look at the present and the future, which is also highlighted by Morawetz as an important element, which allows the therapist to influence the relationship. The background for this paragraph is taken from Judith Beck's book *"Cognitive Therapy"* and Irene Oestrich's book *"Tankens kraft"*.

Practical use of CBT in teaching how to minimize Human Errors

Within the CBT treatment, the structure is essential, both in the session but also throughout the therapy, e.g. the first therapy session will focus on a review of the client's problem, the client's desires, CBT's way of working and so on. This is exactly the same as when planning a teaching session and parallels are clear. For example, if we look at the structuring of CRM/TRM lessons or the construction of King's *"Twelve-Step Lesson Plan"* (p.48-52). Generally, the CBT has three essential pillars (1) education process, (2) self-monitoring, (3) exploring and testing. It is important to understand that this not a static process, but in the treatment you oscillate between all three pillars. In this process, there is also a constantly reconciliation, discussion and following up on the target. When you are teaching pilots and ATC, it is essential that you as a teacher at all times are aware of and constantly draw parallels to their reality, as King highlights (p.47-52), that means transferring a general understanding of how humans function into a situation relevant for pilots and/or ATC, and furthermore include the understanding of how, when registering that we are making an error, this will be expressed and what signals we receive from ourselves before or while making the error. A great advantage in teaching or in group therapy is the ability to use the group dynamic regarding the way of thinking and thereby create reflections inside every single individual.

One of the main pillars of the CBT structure is psycho-education. The personnel, who have had no previous education within this subject, will first receive a presentation of the cognitive thought, in this the Cognitive Diamond and how thoughts, feelings, body and behavior are linked together. The presentation will be followed by an exercise where the students are given approximately 30 different statements. The students are instructed to determine whether the statements are thoughts, feelings, body expressions or behavior (exploring and testing). Afterwards all statements are discussed one after one, increasing their understanding of the concept and theory. The next step a presentation of the relevance of how to understand themselves and the world surrounding them, the influence of negative thoughts, but also that a thought is just a thought. A thought might be right, it might be wrong, but you can test thoughts, and thoughts can change. Next step is an exercise (described in the next section), in which the students fill out diagrams describing how they react. This last part is the first step of providing the students tools to self-monitoring. During this process they are introduced to the chain analysis, which is a method to understand and modify unwanted incidents. This provides the students a concrete and practical method to monitor themselves and test their thoughts and eventually discover alternative strategies.

An important process to make students aware of when they need to obtain a new behavior is the educational process. This process consists of different steps and the students should be aware of what they must pay attention to in the steps. When talking to students, it seems that they are giving too much attention to the phase of flight/live traffic control (ATC) and almost forget what they need to work on between sessions. Figure two is shown to the students while teaching. The session step indicates the time of flying or being in a live traffic situation. When teaching, each phase is examined separately and psycho-educating is done on what to be aware of, followed by discussion and transfer of the knowledge into relevant situations, e.g. a discussion of how to maximize the gain of an instructor during a debriefing. In this phase, focus is on questioning techniques, in-depth questions that lead to greater understanding, but also on making the students aware of and focus on their own reaction in relation to the information that they receive from the instructor, again to increase the awareness of their own reactions. Next phase is reflection and analysis. In teaching, we show two different shooting boards (see figure 3), and ask which one they would wish they had done. Every time, almost everyone say the right one. Then the points attained from shooting are shown, the left receives 88 points and the right 61 points. The instructor informs that students have lost the game, and quickly the discussion sets off with the argument that the right shooting board is easier to correct, e.g. sight is not set correctly. In this phase, focus is on getting a discussion of whether there is a pattern versus a coincidence, how to find out when to be extra vigilant, identify alternative behavior patterns and thoughts. In other words, learn to reflect after a debriefing and how to analyze what needs to be developed before the next session. Based on the results of the analysis, next phase is to work on the desired behavior. It may be that the students need to understand more theory, practice in a simulator, have mental training, do visualization exercises and step for step training. The last phase before the new session is briefing, focusing on expectations and what they must focus on, what they discussed in the last debriefing, etc.

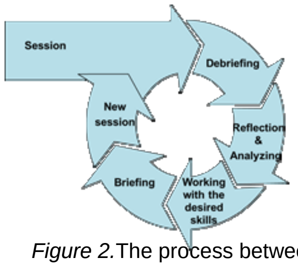


Figure 2. The process between the flights/live traffic for ATC.

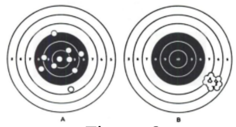


Figure 3.

As described, a large part of the cognitive understanding is related to how you think and if the result of the way the person thinks is destructive, it is important to find alternative ways to think about it, that is, too reformulate your thoughts. An example, when ATC identify a conflict on the radar screen (the stress level is high), accidentally they will tell the pilot to climb to a new level, even though this creates a new conflict. What's interesting, is that the ATC usually already has observed the conflict and is aware of it, and yet he still tells the pilot to climb to the new level and create a new conflict. From a neurological theory called the Serial Position Effect, we know that people often remember the first and the last thing, in other words, it is what's in between you forget (Gade, p 218). When asking ATC's how they think in situations like this, many describe that they scan the radar screen, that they often prior to the incident have had a conflict, and when observing the new situation they think "I must not say, climb to flight level 300". Due to the rapid change of situation, they end up doing the exact opposite and tell the flight "climb to flight level 300" creating a conflict. Instead, the focus should be on the desired outcome, thus to think of what to do, e.g. telling the flight to climb to flight level 280. Attention should be on what to do and not on what to avoid. When you are working to change a behavior, it is important to make students aware of when and how to train the new behavior.

As I described in the introduction it happens that students freeze when they make a mistake (often a perfectionist personality) and the error comes to rummage in such a degree that they do not react or act on the error affecting the rest of the flight. A classic statement is that "an error leads to another error", so the importance of breaking the pattern is extremely high. It may also be that the student keeps thinking about *not* making mistakes, leading to actually making mistakes. In this situation you can work both long-term on the student's personality and/or focusing on the situation here and now. This may require supervision of the instructors on how to manage the student and sometimes coaching of the student herself. In teaching, we usually discuss what our experience tells us, that everyone makes mistakes and that it is normal to be annoyed when you make a mistake (normalization), but also to learn from their mistakes and that's it is on the flying school they can make mistakes while they have an instructor at their side. In the discussion all agree that it is unsuitable to have an over-reliance on errors during the flight. This is exemplified by an exercise among the students – the instruction is: "*You may not think about what I say - a pink elephant*". What happens – well, you think about a pink elephant. This is where the theory and understanding of the term acceptance enters. Simplified, the idea is, the faster you accept it, the faster you find a solution and therefore the faster you can move forward.

Example of an Exercise

One of the exercises used in teaching is to give students an understanding of how thought, emotion, body and behavior works. The exercise is used relatively early in the process to give them a practical understanding of the model. Additional reason for this exercise is to give students the awareness that they might react differently to the same situation occurring several times, even though it is in fact the same situation, additionally how the reaction is different.

The first step is to gather the group in a circle, the students receives a balloon – often laughter and curiosity about what will happen arises already, being a positive sign, that you have their attention. This is a clear parallel to both Huddle and King's descriptions of what is important in attaining a positive result. The instructions follow to inflate the balloon, tie it and hold it between your hands at stomach height and close your eyes. They cannot open their eyes or say anything, until the instructor lets them. The situation changed rapidly from laughter to silence and a certain amount of tension. The instructor chooses a balloon and blasts it with a needle. The first reaction of the students is often a pair of open eyes, a few laughs, a few tense facial expressions, and others states afterwards that they started listening more carefully when the instructor entered the circle. The students are now asked to take their seats again and are handed a scheme to write down their experiences regarding the situation, the thoughts they had regarding of the situation, their emotions and how strong these emotions were on a scale of one to ten, and how their body responded when the balloon bursted. When everyone is finished filling out the scheme the instructor ask how they experienced filling out the scheme, if there were anything they were uncertain of, etc. Afterwards, the students are again asked to gather in a circle and the exercise is repeated including the filling of evaluations schemes. The instructor observes the group and notes in his memory differences in group behavior between the first and the second sequence.

The next step is discussion; the instructor starts by asking if someone wants to tell what she wrote, both the first and the second time. The instructor goes through a couple of experiences, and focuses on the differences, how they were expressed, whether they tried to have a strategy in the second sequence, if it was easier to fill in the scheme second time, etc. The last step is to transform the exercise to "reality", in other words working on transferring the experiences to their reality and discussing what this exercise reflects, what can be transferred to the flight deck, how it can be understood in relation to errors, how can we use it, what their answers reflect, what they should work on and so on.

Generally you often see the reactions dividing into three groups, one group expressing that they knew what would happen the second time and therefore were not that uncertain, the second group expressing that they do not experience the big difference between the two sequences and the last group expressing that they became more nervous the second time, because now they knew what would happen and they were certain their balloon would be picked. The instructor relates every reaction to making errors and what it would mean in "real life" with these kinds of reaction patterns, and what to work with in every reaction pattern. Experience shows that it is the latter group that needs to work the hardest to manage and develop a more constructive way to pass the education but also to respond to the errors they make.

The experiences with this exercise are very positive. This exercise (exploring and testing) gives the students tools and a scheme to learning how to register their own reactions (self-monitoring). Furthermore, the strength is related to providing an understanding of how a perception can be changed even though it is the same basic situation.

Evaluation of the implementation of cognitive therapy methods in teaching

Three different groups of personnel have been subject of the evaluation of the education of cognitive psychology; these groups are ATC students, students from the Royal Danish Air Force Flying School and employees within the Royal Danish Air Force (RDAF). All courses within cognitive psychology in the RDAF have been evaluated. Therefore, the results at hand will provide insight into the relevance of cognitive psychology in all relevant functions within the Air Force, both from a student point of view as well from current employees within the RDAF, from Pilots and Air Traffic Controllers to Technical Personnel and Mission Planner.

Timing of the cognitive psychology course has been planned according to the elements and flow of the overall education of the students and the employees. From this viewpoint a 3 day course was planned just prior to startup of On Job Training for the ATC students providing them instruments relevant when turning to the more “real” part of the education. At the Royal Danish Flying School the training within cognitive psychology was planned as a separate course and for the employees within the RDAF the course was implemented as part of continuing education. The training material focused on the same issues within cognitive psychology, but was of course targeted to the various groups making the education as targeted and relevant as possible.

Immediately after completion of the course, participants evaluated its utilization, keeping the evaluation up to date and making sure it would reflect the actual experience of the course. Furthermore, the evaluation process was separated from the education, securing unbiased results as the teacher had no part in the evaluation process. Evaluation was either conducted as an online survey or by using handout evaluation schemes.

Education within the field of cognitive psychology was initiated in 2009, however, it was not until 2010 an organized evaluation of education was put in place. This paper includes all evaluation since 2010.

All questionnaires used for the evaluation was divided into two parts. The first part of the questionnaire included scales providing quantitative measures of the assessment of the course. The scales covered academic content, relevance as well as skills of the educator. The second part provided an opportunity to deepen the evaluation with personal comments.

Results

The evaluation of the cognitive psychology course included dimensions related to the fundamental elements of the cognitive psychology and of course evaluation of the perceived professional gain from attending the course. The results are divided into the three different groups of personnel.

Psychological Factors in Aviation - ATC students			
Assessment of following dimensions (share):			
	:-(	:-|	:-)
Overall assesment/ Professional Gain	0%	9%	91%
	Not relevant		Relevant
Professional Relevance	0%		100%

Looking across all results it is evident that the education pays off. All three groups of personnel assess the education very positively. Looking at the results among students, table 1 and 2, all ATC students find the course professionally relevant; Furthermore, 10 out of 11 students have a positive assessment of the professional gain from the course. Likewise, the ranks from the students of the Royal Danish Flying School all range from ‘Good’ to ‘Excellent’. Especially, the evaluated Professional Relevance and Professional Gain are of particular interest when evaluating the eligibility of the course. Both dimensions rank high underlining the appraisal of the elements of cognitive psychology implemented during training.

Human Factors, Psychological Factors in Aviation - RDAF Flying School							
	2012						
Assessment of following dimensions (share):	1: Very Poor	2: Poor	3: Good	4: Very Good	5: Excellent	*Very good & Excellent - Share*	*Mean*
Professional Relevance	0%	0%	44%	33%	22%	*56%*	*3,8*
Overall assesment/ professional Gain	0%	0%	44%	33%	22%	*56%*	*3,8*
Teachers Academic Level	0%	0%	44%	22%	33%	*56%*	*3,9*
Teachers Motivation	0%	0%	44%	44%	11%	*56%*	*3,7*
Teachers Presentation	0%	0%	56%	33%	11%	*44%*	*3,6*
Relevance of Teaching Me	0%	0%	78%	11%	11%	22%	3,3
Teaching Materials	0%	0%	44%	33%	22%	56%	3,8

In table 3 we have the employees within the RDAF, where the course was implemented as part of continuing education. Even here the assessment shows the strong relevance of cognitive psychology. In 2010 the Professional Relevance was evaluated extremely well with more than 3 out of 4 ranking it ‘Very good’ or even ‘Excellent’. In 2012 the evaluation decreased a bit with half the participants ranking it ‘Very good’ or ‘Excellent’. However, we must keep in mind that is still a very strong assessment. Furthermore in 2012 the group of participants was much more diversified than the previous two years, making it more difficult to embrace all functions equally well.

Looking further into these results we see that the part of the course showing the most positive increase are the dimensions related to planning and organization. From the 2010 results it became clear that the participants needed more time devoted to cognitive psychology in order to get the full value of the principles. And since the relevance and gain from the course was very satisfying, the decision was made to increase the number of lessons. Today, the number of hours has doubled compared to 2010.

To sum up, the results clearly indicates the relevance of cognitive psychology within aviation. However, though all participants have evaluated the course, we need to keep in mind that the number of evaluations are still limited and should be assessed from a qualitative point of view.

Human Factors, Psychological Factors in Aviation - EMPLYOEES - MEANS			
Assessment of following dimensions (share):	2010	2011	2012
Professional Relevance	3,9	3,9	3,6
Professional Gain	3,7	3,5	3,7
Teachers Presentation	3,7	4,0	4,0
Teachers Academic Level	3,9	3,9	4,0
Teaching Materials	2,9	3,4	3,3
Assessment of following dimensions (share):	2010	2011	2012
Academic Level	3,4	3,4	3,3
Time in relation to Curriculum	2,2	3,0	2,9

Human Factors, Psychological Factors in Aviation - EMPLYOEES						
	2010					
Assessment of following dimensions (share):	1: Very Poor	2: Poor	3: Good	4: Very Good	5: Excellent	*Share of 'Very good' & 'Excellent'*
Professional Relevance	0%	0%	22%	61%	17%	*78%*
Professional Gain	0%	6%	33%	50%	11%	*61%*
Teachers Presentation	0%	0%	44%	44%	11%	*56%*
Teachers Academic Level	0%	0%	22%	61%	17%	*78%*
Teaching Materials	0%	11%	67%	17%	0%	*17%*
Assessment of following dimensions (share):	1: Not at all adequate	2: Not adequate	3: Adequate	4: Not adequate	5: Not at all adequate	*Share of 'Adequate'*
Academic Level	0%	0%	56%	44%	0%	*56%*
Time in relation to Curriculum	17%	50%	33%	0%	0%	*33%*
	2011					
Assessment of following dimensions (share):	1: Very Poor	2: Poor	3: Good	4: Very Good	5: Excellent	*Share of 'Very good' & 'Excellent'*
Professional Relevance	0%	0%	26%	61%	13%	*74%*
Professional Gain	0%	0%	57%	39%	4%	*43%*
Teachers Presentation	0%	4%	22%	43%	30%	*74%*
Teachers Academic Level	0%	0%	22%	65%	13%	*78%*
Teaching Materials	0%	0%	64%	36%	0%	*36%*
Assessment of following dimensions (share):	1: Very Poor	2: Poor	3: Good	4: Very Good	5: Excellent	*Share of Adequate*
Academic Level	0%	0%	57%	43%	0%	*57%*
Time in relation to Curriculum	0%	4%	96%	0%	0%	*96%*
	2012					
Assessment of following dimensions (share):	1: Very Poor	2: Poor	3: Good	4: Very Good	5: Excellent	*Share of 'Very good' & 'Excellent'*
Professional Relevance	0%	0%	50%	41%	9%	*50%*
Professional Gain	0%	0%	45%	36%	18%	*55%*
Teachers Presentation	0%	0%	18%	64%	18%	*82%*
Teachers Academic Level	0%	0%	27%	50%	23%	*73%*
Teaching Materials	0%	0%	82%	9%	9%	*18%*
Assessment of following dimensions (share):	1: Very Poor	2: Poor	3: Good	4: Very Good	5: Excellent	*Share of Adequate*
Academic Level	0%	0%	73%	27%	0%	*73%*
Time in relation to Curriculum	0%	9%	91%	0%	0%	*91%*

Discussion

In this article, I have presented the views of both Hubble and King regarding factors that must be in place in therapy and education to achieve a positive result. In the evaluation schemes this can be expressed as the academic level of the teacher, his motivation and his presentation skills. All in all, 71% states that the academic level was ‘Very Good’/‘Excellent’. Similar results are seen regarding presentation skills (64%) and motivation (56%). No one evaluated the dimensions less than ‘Good’. Hence, one can conclude that teaching has met the expectations and generated interest.

It has not been possible to determine whether the education alone has lead to lesser errors or to detecting errors faster. This is due to terms of resources and the research aspects. A general problem is how to study evidence that the education actually has a positive outcome in relation to minimize errors. Instead we can turn to look at the evaluation of teaching and what people think about the relevance and gain from it. All in all, 71% states that the education is ‘Very Good’/’Excellent’ in relation to relevance, and the gain from attending the course is likewise by 61% stated to be ‘Very Good’/’Excellent’. Almost no one evaluated the dimensions less than ‘Good’ (except for one person in gain). Furthermore, verbatims from the students underline these results, providing clear signals that the cognitive theory is relevant to this audience. These are some of the verbatims from the students: *"It was great getting "tools" from the psychologist that I can use in my work every day.", "A lot of useful hints and tools", "Wonderful with tools that you can bring home with you and use, and that we have the opportunity to test and train the methods while help is at hand", “A lot of things have been examined, and if things are done like this, I’m sure, it would make everyday life a whole lot easier. I think that we all take a small part of it home, and are excited to see results.”*

Perspective and Conclusion#

Results from the evidence-based research identifies some general factors that should be in place for a positive outcome of treatment. The teaching should certainly include these general factors to ensure contact, motivation and the good relationship between the teacher and the students. Education is of course also about learning why it is not enough to just look at general factors, but it is important to look at the evidence-based research within the different treatment methods. The treatment that is mostly emphasized at the moment is the cognitive behavioral therapy. In the article, I have shown the underlying theory and the practical transfer to education for pilots and ATC area in relation to how they understand themselves, by learning the cognitive understanding through the education process, self-monitoring and exploring/testing. The idea is that by increasing understanding you can avoid/reduce/manage errors. As revealed in the result and discussion part, clear indications are given that the education is relevant and with a large gain. With that said, it is important to remember that the article and the evaluation results should be seen as an indication and inspiration to practical use and future follow-up studies and research.

In this article, the focus has been on the cognitive therapy's role in teaching. There are indications of acceptance and visualization techniques received positively by the students, but in the evaluation the focus has been on the overall education. It would be interesting to examine more specifically the concepts or elements of learning to look at the effect. A concept that I think should be explored further in relation to Error Management is the term 'accepting'. In therapy we know that the concept is very important, because, the faster a person can accept an idea or a situation, the faster will he find the solution to the problem and move on. Hence, it would be the interesting if it can be transferred to a cockpit situation, in other words if focus will be aimed faster on solution? CBT has been transferred to many other therapeutic areas, and it can be concluded from these results, that it is also possible successfully transfer CBT to education of pilots and ATC.

References

Beck, S., Judith (1995) Cognitive Therapy. *Guilford Publications.*

Butler, A., Andrew, Chapman, E., Jason, Forman, M. Evan & Beck, T., Aaron (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review,* 26, 17-31.

Free, L., Michael (2007). Cognitive Therapy in Groups. Guidelines and Resources for Practice. (2nd ed.) *John Wiley & Sons, Ltd.*

Gade, Anders (1997). Hjerneprocesser: kognition og neurovidenskab. *Frydenlund Grafisk.*

Hubble, A., Mark, Duncan, L., Barry & Miller, D., Scott (1999). The Heart and Soul of Change: What Works in Therapy. *American Psychological Association*.

King, R (1999). Aerospace Clinical Psychology. *Ashgate Publishing Company*.

Morawetz, D. (2002) What works in therapy? What Australian clients say. http://www.psychology.org.au/Events/Downloads/633704683085518184_What%20Works-4-in%20Therapy-PsychOz-long%20version-1.doc

Oestrich, H., Irene (2000). Tankens kraft. Kognitiv terapi i klinisk praksis. *Dansk Psykologisk Forlag.*

Contact information

Patrik Jonzon
Sergels väg 13a, lgh. 1302
217 57 Malmö
Sweden
Phone: +46709606632
E-mail: p_jonzon@hotmail.com or patrik@mil.dk

TRAINING FOR SITUATION AWARENESS

Michael T. Hertz
Progeny Systems Corporation
Manassas, VA 20110
Norman R. Hertz
Progeny Systems Corporation
Manassas, VA 20110

Situation awareness (SA) is not fixed but is malleable and can be enhanced by training. The operating premise is that SA is measurable given the many theories of SA. This paper identifies the various methods and approaches that have proven effective in training for greater degrees of SA. The loss of ,or insufficient SA is not viewed as inattention or lack of focus but instead, it is very likely that the person lacked the skill set to perform the job. The amount of training necessary for the person to effectively perform the work is helpful if it reduces the workload and informs decision making. A person is set up for failure if one of the requirements for the job requires a high degree of situation awareness without a sufficient skill set. This paper will present some suggested methodologies for enhancing SA to overcome the notion that SA is immutable.

Obviously the topic “Training for Situation Awareness” is not new. Endsley and Robertson (2000) were published in a book chapter that had the same title as this paper with the added words “Individual and Teams.” There is no intent to model this paper after the earlier publication but will approach the topic from a point of view that there is no preferred model for measuring SA and errors of judgment occur without association with SA. For example, error chains are formed when errors of judgment are made just from a lack of ability.

Situation awareness “means that the pilot has an integrated understanding of the factors that will contribute to the safe flying of the aircraft under normal or near-normal conditions. The broader this knowledge is, the greater the degree of situational awareness” (Regal, Rogers, and Boucek, 1988, p. 65) as cited in Adams, Tenney, and Pew (1995). Adams, Tenney, and Pew (1995) states, “this definition stresses that successful realization of these processes depend on the prior knowledge with which the pilot meets such data.” They report that the definition also stresses”that the ideal body of prior knowledge is prodigious in depth and breadth, as it includes not only up-to-the-moment understanding of current flight and aircraft status but also the background knowledge that lend familiarity to or permits understanding of any datum that could arrive in terms of the full range of situations, implications, and response options that go with it (pages 86-87).”

It is possible to increase situation awareness by training which provides prior knowledge. Situation awareness is not based upon native intelligence but instead is based upon knowledge that is developed during training. In the literature, situation awareness is sometimes treated as a personality trait such as extraversion. Personality traits by definition are enduring and unchanging. As long as situation awareness is thought of as a personality trait, the understanding of situation awareness and situational awareness training will not advance.

If situation awareness was an enduring personality trait then it would not be necessary to provide training. A person could be administered a "situation awareness" test and the results used for job placement. In practice, situation awareness is treated as trainable for there is no other reason why previous jobs experience or professional supervision would be required as conditions of employment. Industry treats training and previous experiences as risk mitigation in the situational awareness domain.

Training Perspectives

Training situational awareness can work in this fashion: Through the development of schemata (long-term memory) through training, the operator does not need to attend to every detail of the environment to have a reasonably complete assessment of the situation. This model of SA predicts that the relationship and quality of SA is dependent on the completeness of the knowledge the pilot has stored in long-term memory and the operator's working memory capacity .

> "The very definition of SA implies that human performance in any task cannot improve unless the trainee begins to develop a comprehensive body of domain specific knowledge and a detailed understanding of how the knowledge should be used to improve task performance." (Vincenzi, Hays, and Seamon, 2000, page 364).

SA is of interest to pilots—both aviation and maritime, power plant operators, and process control operators because their performance is affected by the amount of SA present. The delta between good and bad or effective and ineffective decisions is based on a proper understanding of the current system. Researchers design interventions to improve SA and find that performance is improved without an increase in SA (Brooks, Switzer, & Gugerty, 2003). These findings suggest that SA training affects performance without a concomitant increase in measurable SA. These findings may mean that the SA measures are insufficiently robust or that the construct of SA may not be independent from performance. Training for SA could very well focus on increasing job skills.

The following underlying competencies are considered as potential candidates for training (Banbury, Dudfield, Horman (2004, page 80) :

1. "to think ahead to future phases of the flight, instead of simply noticing events, in order to maintain SA
2. to perceive loss of SA, both of their own and of others, and to act on that knowledge
3. to re-evaluate criticality decisions by seeking data to disprove, rather than confirm, the current course of action
4. to balance workload, both manual and cognitive, between crew-members effectively."

The items will be addressed in order.

- Item 1—training could be designed such that it is standard operating procedure for person to visualize the flight in its entirety prior to beginning the flight or to visualize the entire process control activities. The visualization will prepare persons to be able to predict and plan for alternative actions prior to actual occurrence.
- Item 2—persons should be trained to identify the behavioral characteristics of themselves and others when loss of situation awareness occurs. Moreover, they should be prepared to act upon the behavioral indicators of the loss.
- Item 3—the critical approach is to disconfirm the decisions rather than to confirm because of the prevalence of confirmation bias. The process of disconfirmation is a standard procedure in philosophical approach to advancing science.
- Item—Excessive workload will cause the person to miss key indicators of events that begins the development of an error chain.

Bolstad, Endsley, Howell, and Costello (2002) evaluated two training modules—preflight training and contingency planning. "Their findings were that the modules were somewhat successful for improving SA and the pilots found them informative and useful (page 25)". Similar content to preflight training and contingency planning were identified earlier in this paper as domains in which training would assist in enhancing SA.

Endsley and Garland (2000) provided training recommendations for improving SA. They reported that good task management strategies appeared to be critical for dealing with task interruptions and distractions. The development of comprehension was another area in which training would be helpful. The third area in which training would be of benefit would be in planning in order to anticipate future events. A final area to provide training would be to assist persons in performing their own situation assessments.

Fowlkes, Merket, and Oser (2000) state that SA is vaguely defined and there is little prescriptive guidance available for how to train for SA. They suggested that behavioral indicators can be used to infer whether crews note relevant information. They further stated that behavioral statements lend themselves to the development for training objectives and ultimately to the development of assessment tools.

Sethumadhaven (2011) stated that only when individuals make accurate meta-cognitive judgments about SA can they adopt better monitoring strategies and be equipped to handle automation failure. The results from their study suggest that controllers who had better confidence in their SA tended to have better SA and those with better SA were faster in responding to automation failure. It is possible that meta-SA training programs can be used to improve comprehension of operators in dynamic situations by helping operators develop better monitoring and control strategies.

Training Content

Comprehensive body of knowledge
Plan and think ahead (be out front of the aircraft or ship)
Perception of loss of SA
Disconfirm critical decisions
Distribute workload
Preflight training
Contingency planning
Task management
Development of comprehension
Planning to anticipate future events
Perform self assessment of SA
Behavioral Application of training objectives
Metacognition

Based on the training content identified in the literature, there is no magic bullet that points the way directly to increasing SA. SA is a complex phenomenon that is used to explain deficient performance while the literature is saying that SA is not performance. If not, performance, than what? Deficient performance can be explained by lack of training, lack of cognitive or psychomotor skills, motivation, or other variables. One of the reasons that the SA research has generated the volumes of research has been the failure of the research to approach SA as a unitary construct.

How has the development of measures of SA improved performance? It would more productive if training was focused on the components of SA such as, planning, the process of disconfirming theories, task management, and metacognition.

The proposal advanced in this paper is that the focus of situation training should be in individual components instead of the end product of SA. SA should be decomposed on the constructs that predict performance and hence SA.

Acknowledgements

The opinions and cited research in this paper do not reflect the views of Progeny Systems Corporation.

References

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors, 37*, 85-104. doi: 10.1518/001872095779049462

Banbury, S., Dudfield, H., & Hörman, H –J. (2004). Development of novel measures to assess the effectiveness of commercial airline pilot situation awareness training. In *Proceedings* *of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 80-84) Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193120404800118

Bolstad, C. A., Endsley, M. R., Howell, C., D.,& Costello, A. M. (2002). General aviation pilot training for situation awareness: An evaluation.). In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 21-25) Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193120204600105.

Brooks, J. O., Switzer, F. S. & Gugerty, L. (2003). Effects of situation awareness training on novice process control operators. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 606-609) Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193120304700375.

Endsley, M. R. & Robertson, L. M. (2000). Training for situation awareness in individuals and teams. In M. Endsley and D. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 349-365). Mahwah, NJ: Lawrence Earlbaum Associates.

Endsley, M. R. & Garland, D. J. (2000). Pilot situation awareness training in general aviation. In *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* (pp. 357-360) Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193120004401107.

Fowlkes, J. E., Merket, D. C., Oser, R. L. (2000). Transitioning SA theory and research into practical training guidance: A case study. In *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* (pp. 419-422): Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/154193120004401124 Correct Reference: Proceedings of the IAE 2000/HFES 2000 Congress 2-419

Regal, D. M., Rogers, W. H., Boucek, G. P. (1988). Situation awareness in the commercial flight deck: Definition, measurement and enhancement. In *Proceedings of the Seventh Aerospace Behavioral Technology Conference and Exposition* (pp 65-69). Warrendale, PA: Society of Automotive Engineers.

Sethumadhaven, A. (2011). Knowing what you know: The role of meta-situation awareness in predicting situation awareness. . In *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting* (pp. 360-364): Santa Monica, CA: Human Factors and Ergonomics Society. doi: 10.1177/1071181311551074

Vincenzi, D. A., Hays, R. T., & Seamon, A. G. (2000). Measuring situation awareness in training systems: A multivariate approach. In *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* (pp. 361-364): Santa Monica, CA: Human Factors and Ergonomics Society. 361 DOI: 10.1177/154193120004401108 Correct Reference Proceedings of the IAE 2000/HFES 2000 Congress 2-361

A Proposal to Reduce Unsafe Aviation Maintenance Task Handovers with a Virtual Training Solution

Abstract

Kevin Gildea, Wayne Shebilske, Ganesh Alakke, and Shruti Narakesari

Safety researchers have documented maintenance errors happen when outgoing crews inadequately communicate to incoming crews at task/shift handover. Poor task handovers are especially likely for low-frequency problems, when complacency is present, or with new technicians. Highly experienced maintenance personnel may have prevented the poor handover. Backup safety procedures catch most errors, but correcting the remaining increases risks to safety, such as those not caught until after the plane is rolled out of the hanger. The cumulative effect of handover errors increases the exposure to additional threats and errors during reworks and replacement of aircraft on the line among others. This poster introduces a virtual training solution adapted to reduce aviation maintenance handover errors. The solution pairs experienced personnel with less experienced personnel. It also increases virtual handover frequency for infrequent problems so that less experienced personnel are prepared to make successful handovers even for problems they have never experienced on the job.

Introduction

The efficiency and efficacy of aviation maintenance are critical to the safety of air operations. With the current safety record as a backdrop, the aviation community is focusing on reducing the risks of injury and damage. Example methods employed that have aided in the current climate of safety are procedures and practices that capture and correct errors before they can lead to an accident. Even when no injury results exposure to additional threats and errors can increase. In any domain where two or more people work sequentially on tasks, there is a challenge of properly informing those who will take over the task concerning the state of the operation (Parke, Hobbs, & Kanki, 2011; Parke & Kanki, 2008). Although some tasks may be routine, there are non-routine cases when the next shift arrives or the task is at a point to be transferred to the next maintainer and requires an explanation to those receiving the task.

An instance where this was particularly apparent was the cause of the crash of an Embraer 120 that suffered a structural breakup in flight and crashed near Eagle Lake, Texas (NTSB, 1992). During the investigation, it was found that a second shift mechanic removed screws from the tops of both the left and right horizontal stabilizer leading edge assemblies. This was done in the course of removal and replacement of the deicing boots. Several factors interfered with the third shift fully completing the maintenance on the deicing boots. However, a critical issue was a lack of communication concerning the removal of the screws from the tops of both the left and right boots. The Embraer 120 is a T-tail aircraft with the top of the horizontal stabilizer not being visible from the ground. Without complete information on the state of the aircraft, the right boot was removed and a new boot bonded and secured. However, the aircraft was placed into service without the top screws.

Given the complexities of the aviation maintenance environment with numerous complex tasks, it is an ongoing challenge to avoid such oversights (Loukopoulos, Dismukes, & Barshi, 2009). Maintenance teams are faced with communication challenges throughout the shift, particularly during task handovers. Even with robust procedures and substantial levels of training, opportunities for shift handover errors are ever-present. The required communication skills present unique challenges for novice members of maintenance teams, thus opening opportunities for the application of virtual training methodologies (Shebilske, Gildea, Freeman, & Levchuk, 2009; Shebilske, Goettl, Jordan, & Day, 1999).

Proposed Training Solution

Aviation maintenance environments place demands on team members to engage in frequent, proactive exchanges of potentially complex information (Jiang, Master, Kelkar, & Gramopadhye, 2002). The time pressure and complexity increase the criticality of selecting what information to convey, when, to whom, and in what format. Training communication skills for such time-compressed environments is complicated by the time compression itself. One of the most effective ways to teach communication skills is through demonstration, role playing, and timely, corrective feedback. A virtual training approach is proposed that provides the opportunity for experienced developers to model effective communications, guide newer employees, and provide feedback. The training environment is designed to provide opportunities to practice identifying critical information, determining who needs the information, selecting the appropriate time to push or pull the information, and using various information transfer methods. This training approach supports cataloging lessons learned for use in future training scenarios.

Previous research conducted with military teams in time-pressured, complex, operational environments indicates that team training in virtual environments can yield significant gains in performance (Gildea, Schneider, & Shebilske, 2007; Shebilske, Gildea, Ieoger, Volz, & Yen, 2005). Similar results have been found with training for aviation maintenance teams (Kraus & Gramopadhye, 2001). Specifically, this past research has supported the concepts embodied in observational learning and the benefits of modeling skills by experts (Shebilske, Jordan, Goettl, & Paulus, 1998) combined with practice followed by feedback (Shebilske, Gildea, Freeman, &

Levchuk, 2009). The strengths of distributed training in virtual environments include: (1) the ability for geographically separated individuals to practice in concert, (2) an environment conducive to exploring the problem space with the latitude to learn from errors, (3) the capability of presenting critical incidents and lessons learned in a shorter amount of time than required to naturally experience such situations in the real world, and (4) the opportunity to practice in a virtual context that is representative of the work environment, thus supporting transfer of training.

The proposed virtual training environment is based on the Neverwinter Nights™ gaming engine, which has been used by a number of organizations for training development and delivery. The version that will be modified for this effort has been developed by the Culture and Cognition Lab at Wright State University to support multicultural training for teams. This platform supports the capture and analysis of expert behaviors conveyed during the training. This information can lead to a systemization of the knowledge for implementation in future iterations of the training. This effectively leads to the training suite serving as a research vehicle to enhance understanding of team interactions.

The authors have previously adapted virtual training environments to train teamwork skills for time-critical asset allocation and multinational team training (Shebilske, Levchuk, Freeman, & Gildea, 2010). These training environments provide the foundation for a means to expose maintenance teams to virtual training for communication or operational interactions. Based on prior experience analyzing new domains, defining requirements, and developing the training system, the authors will extend the extant training environment to maintenance teams.

Operational Benefits

As in the brief example of engine failures, in general, noticing and correcting errors earlier results in less exposure to risk for the passengers, crew, and associated personnel. If a maintenance team effectively manages a handover verbally with supporting written turnover sheets, the risks are reduced. Any errors in the handover can potentially lead to rapidly escalating threats as the time progresses before the error is noticed and corrected. The progression through installation, close up, signoff, returning to the line, etc. at each succeeding step adds additional complexities and time to rectify the error.

Although there are numerous tasks that could benefit from the implementation of such a training approach, we focus on the critical boundaries between team members where information flow is often informal and perhaps semistructured. Those who have been on maintenance teams for extended periods tend to build an understanding of what information is required and at what time by others on the team. When communication is effective delays and costs decrease. Such problems are more likely to occur while training and integrating junior mechanics into work flows. Managers could increase safety and shorten delivery times by more effectively transferring the schemas and complex skill sets to new employees.

By providing a training countermeasure that allows mechanics to practice handovers, with the presentation of lessons learned and interaction with highly experienced mechanics, the probabilities of ineffective handovers should be reduced, which will improve safety and lower costs. A collaborative, virtual training tool will provide the framework and vehicle for training delivery and for cataloging lessons learned to further improve ongoing training.

References

Gildea, K. M., Schneider, T. R., & Shebilske, W. L. (2007). Stress appraisals and training performance on a complex laboratory task. *Human Factors, 49,* 745-758.

Jiang, X., Master, R., Kelkar, K., & Gramopadhye, A. K. (2002). Task analysis of shift change activity in aviation maintenance environment: Methods and findings. Retrieved from http://www.hf.faa.gov/opsmanual/assets/pdfs/Analysis_of_Shift_Change.pdf

Kraus, D. C. & Gramopadhye, A. K. (2001). Effect of team training on aircraft maintenance technicians: Computer-based training versus instructor-based training. *International Journal of Industrial Ergonomics*, *27*, 141-157.

Loukia D. Loukopoulos, L. D., Dismukes, R. K., & Barshi, I. (2009). The Multitasking Myth. Burlington, Vermont: Ashgate.

National Transportation Safety Board. (1992). Aircraft Accident Report, Britt Airways, Inc., d/b/a, Continental Express Flight 2574, In-Flight Structural Breakup, EMB-120RT, N33701. (NTSB/AAR-92/04 PB92-910405). Retrieved from http://www.airdisaster.com/reports/ntsb/AAR92-04.pdf

Parke, B., Hobbs, A., Kanki, B. (2011). Passing the baton: An experimental study of shift handover. Retrieved from http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/ 20110008267_2011008026.pdf

Parke, B. & Kanki, B. (2008). Best practices in shift turnovers: Implications for reducing aviation maintenance turnover errors as revealed in ASRS Reports. *International Journal of Aviation Psychology, 18,* 72 – 85.

Shebilske. W., Gildea, K., Freeman, J., & Levchuk, G. (2009). Optimizing instructional strategies: A benchmarked experiential system for training (BEST). *Theoretical Issues in Ergonomics Science, 10,* 267-278.

Shebilske, W., Gildea, K., Ieoger, T., Volz, R., & Yen, J. (2005). Agent-based training of distributed command and control teams. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 2164-2168). Santa Monica, CA: Human Factors and Ergonomics Society.

Shebilske, W. L., Goettl, B. P., Jordan, J. A., & Day, E. A. (1999). Cognitive and social influences in training teams for complex skills. *Journal of Experimental Psychology: Applied, 5,* 227-249.

Shebilske, W. L., Jordan, J. A., Goettl, B. P., & Paulus, L. E. (1998). Observation versus hands-on practice of complex skills in dyadic, triadic, and tetradic training-teams. *Human Factors, 40,* 525-540.

Shebilske, W., Levchuk, G., Freeman, J., & Gildea, K. (2010). A team training paradigm for better combat identification. In D. Andrews, R. Herz, & M. Wolf (Eds.), *Human Factors Issues in Combat Identification* (pp. 205-216). Burlington, VT: Ashgate.

EFFECTS OF MOTION CUEING ON AN ATTITUDE RECOVERY TASK

Chris M. Nicholson
Ben Townsend
Andrew Staples
Murray Gamble
Dr. Chris M. Herdman
Centre for Visualization and Simulation (VSIM), Carleton University
Ottawa, Canada

The present research investigated the impact of a motion cueing seat on a simple attitude recovery task. Participants (*n*=10) used a joystick to level an attitude indicator that was tilted at either a magnitude of 20 or 40 degrees from level (left or right). A dynamic motion seat was used to provide either congruent or incongruent motion cues. Relative to a no-motion baseline, incongruent motion caused performance decrements as indexed by an increase in the number of control reversals, a decrease in time level, and more over corrections. Congruent motion cueing did not affect performance on the attitude recovery task.

As long as flight simulators have been in use as a tool for pilot training, the concept of fidelity has been of particular interest within aviation research. A flight simulator's ability to simulate motion is generally associated with its level of fidelity. Numerous studies have been conducted to further the understanding of not only how realistic a simulator can be, but more importantly, how realistic a simulator *needs* to be in order to provide effective training (Grant, Yam, Hosman & Schroeder, 2006; Jian, Shutao & Hongren, 2011). Much of the research on motion cueing has yielded mixed results (e.g., Hays, Jacobs, Prince & Salas, 1992; Meyer, Wong, Timson, Perfect, & White, 2012). In a recent meta-analysis, de Winter, Dodou and Mulder (2012) suggested that providing vestibular motion cues can enhance pilot performance when engaged in tasks involving disturbance motion (e.g., turbulence, engine failure) or in complex tasks that involve maneuver motion (i.e., motion feedback as a consequence of pilot's input). This performance enhancement is generally not observed when maneuver motion is simulated on simple flying tasks.

Much of the research exploring motion cueing in the context of pilot training has been conducted using 6 degree of freedom (DOF) Stewart platforms – most likely due to Federal Aviation Administration (FAA) requirements for high fidelity simulators (Burki-Cohen, Sparko & Bellman, 2011). However, dynamic motion cueing seats have been investigated in an attempt to find a cost effective and practical alternative to full motion simulators (e.g., Pasma, Grant, Gamble, Kruk & Herdman, 2011). There is some evidence that the vestibular motion cues generated by dynamic motion seats can enhance performance on specific simulated flight tasks to the same degree as that provided by full motion simulators (Sutton, Skelton & Holt, 2010; Holt, Schreiber, Duran & Schroeder, 2011). However, further research is required.

The present study was designed to explore the effectiveness of motion cueing on a basic attitude recovery task. The experiment used a horizontal line on a computer monitor as a proxy for an (outside-in) attitude indicator that was capable of tilting left or right through 180°. Participants used a joystick to bring an attitude indicator back to level (0°) after being rotated by 20° or 40° (left or right). On motion trials, a dynamic motion seat provided motion cues that either matched the direction of the symbology (e.g., left rotation of the AI, left motion cue) or did not match (e.g., left rotation of the AI, right motion cue). If the cues provided by the motion seat affect attitude recovery performance, then performance should be enhanced by the congruent motion cues and impaired by incongruent motion cues, relative to when no motion cues are provided.

Method

Participants. Ten adult participants (mean age 30.6 years, *SD* = 8.4) participated in the study. Participants had normal or corrected-to-normal visual acuity. One participant was a military (CF18) pilot with 900 flight hours experience: the data from this participant was undistinguishable from the non-pilot data.

Materials. The attitude indicator symbology consisted of a green line with a small triangular base to simulate a simplified outside-in attitude indicator. The symbology subtended approximately 11° of visual angle and was presented against a black background on a 47-inch LCD monitor with a 60Hz refresh rate. The monitor screen was placed approximately four feet from the participant's viewpoint and was kept stationary. A PC was used to generate symbology and log data.

Motion cues were provided by isolating the seat-pan-tilt and seatback sway motion functions on a 5 DOF dynamic motion seat built by ACME World Wide Ltd. Participants were fastened to the motion seat using lap and shoulder belts with their feet planted flat on the floor. Motion cues consisted of the seat pan tilting left or right at a maximum rate of 200 deg/s accompanied by the seatback sway motion. On motion cueing trials, the seat continued to mimic, congruently or incongruently, the motion of the symbology from initial upset to the end of the trial. The motion seat was driven using the chassis assembly provided by the manufacturer.

Participants controlled the symbology using a Cyborg Evo Inc. non-force feedback joystick placed on a lapboard. Only inputs along the X-axis were read and used to level the symbology. Participants used a hat switch on the joystick to acknowledge the experimental instructions and to advance from one block of trials to the next.

Procedure. Each participant completed 5 blocks of 24 trials, resulting in a total of 120 trials per participant. Motion condition (no motion vs. congruent motion vs. incongruent motion), rotation magnitude (20° vs. 40°), and rotation direction (left vs. right) were crossed within each block of trials resulting in twelve unique trial types. Each trial type was presented twice (randomly) within each block. Prior to the beginning of the experimental trials, participants familiarized themselves with the apparatus and tasks by completing one block of trials. Each trial began with a visual (digit) countdown from 3, followed by the presentation of the rotated attitude indicator. In the two motion conditions, the seat motion cue was delivered concurrently with the presentation of the attitude indicator.

Participants were required to level the symbology as quickly and accurately as possible and to keep the line level until the end of the trial. Trials were five seconds long, starting from the onset of the attitude indicator. Micro control inputs at the start of each trial were muted until the attitude indicator symbology was presented at which point data recording was initiated. Control reversals, over-corrections, time level, and RMSE were recorded at 60 Hz.

Results

Figure 1 summarizes control inputs collapsed across all trials and graphed across time (0-5000 ms). Control inputs were transformed to equalize direction of correct joystick input (range ± 1) across left and right trials. As shown in Figure 1, control inputs were initiated approximately 100 ms sooner in both motion conditions compared to the no-motion condition. Participants also used more forceful control inputs in both motion conditions relative to the no-motion condition. Finally, control inputs in the congruent motion condition yielded a distinct 'step down' pattern as participants approached level (indicated by mark "1" on Figure 1). This pattern of control input suggests that congruent motion affects attitude recovery by inducing a feedback loop between participants' control inputs and the attitude symbology.

Control Reversals. A control reversal was defined as an initial control input that was 0.5° or greater in the opposite direction from level. For example, if the symbology was rotated left and the participant's input rotated it further left (by 0.5° or more), then this was classified as a control reversal. Control reversals were quantified as time spent (in ms) past the 0.5° thresholds. Table 1 shows the number of control reversals, average duration of control reversals, and average magnitude of control reversals varied across motion conditions. As shown in Table 1, there were more control reversals in the incongruent motion cueing condition and these reversals lasted longer and were or greater magnitude as compared to the no-motion condition and congruent motion conditions.

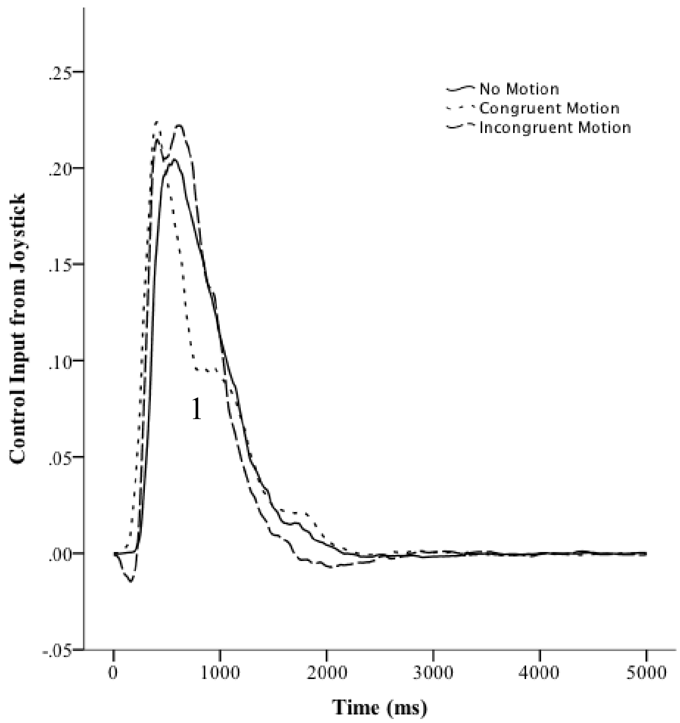


Figure 1. Average control input magnitude across time as a function of motion condition.

Table 1. Descriptive statistics for control reversals.

	No-Motion	Congruent Motion	Incongruent Motion
Number of Reversals	7	15	44
Average Length of Reversal	340ms	309ms	404ms
Average Size of Reversal	3.25°	4.22°	11.37°

Figure 2 shows the amount of time spent in a control reversal averaged across all trials. A 2 (Magnitude: 20° vs. 40°) x 3 (Motion Congruency: no motion vs. congruent motion vs. incongruent motion) repeated measures ANOVA showed a main effect of motion congruency, $F(2, 18) = 3.33$, $MSE = 2591.09$, $p < .059$. Control reversal times were longer in the incongruent motion condition ($M = 44$ ms) than in the no-motion and congruent motion conditions ($M = 6$ ms and $M = 12$ ms, respectively). No other effects were significant.

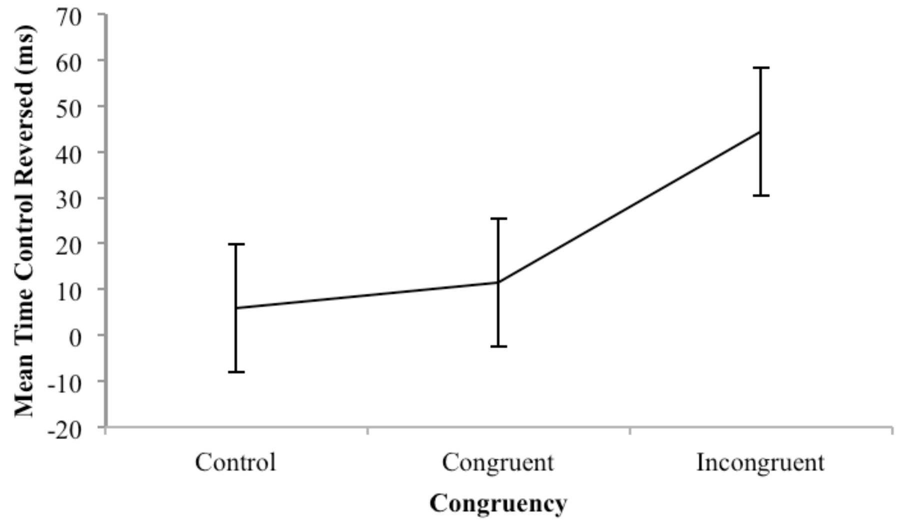


Figure 2. Average control reversal time (ms) as a function of motion congruency and 95% CI [± 15.8ms] (Jarmasz & Hollands, 2009).

Over-corrections. Over-corrections were defined as instances where participants corrected through (past) level by more than 2° and quantified as time spent (in ms) beyond 2° past level. Over-corrections (Figure 3) were analyzed using a 2 (Magnitude: 20° vs. 40°) x 3 (Motion Congruency: no motion vs. congruent motion vs. incongruent motion) repeated measures ANOVA. There was a significant effect of motion congruency, F (2, 18) = 14.27, MSE = 19760.64, $p < .001$, with more time spent over-correcting in the incongruent condition (M = 485 ms) than in the no-motion and congruent motion conditions (M =314 ms and M = 257 ms, respectively). No other effects were significant.

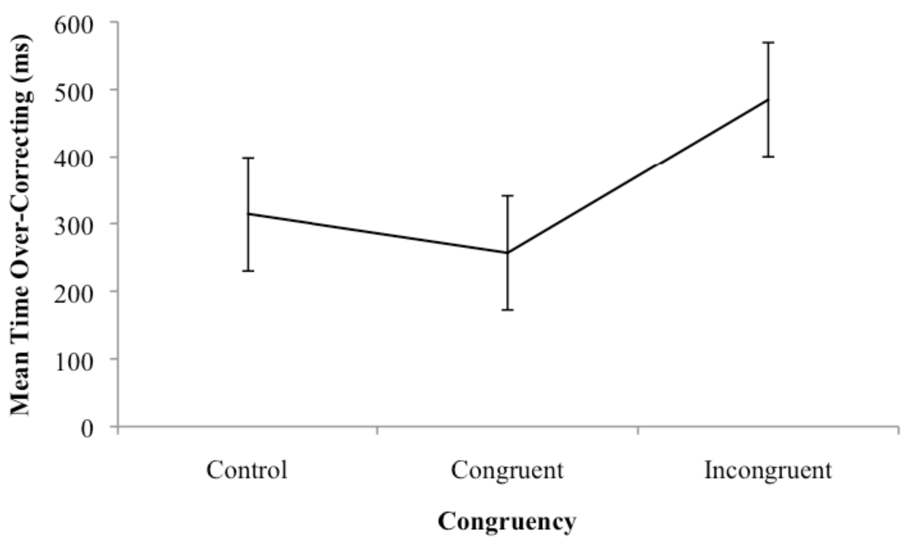


Figure 3. Average over-correction time (ms) as a function of motion congruency and 95% CI [± 84.4ms].

Time Level. Time level was quantified as time spent (in ms) within 2° of level. Time level (Figure 4) was analyzed using a 2 (Magnitude: 20° vs. 40°) x 3 (Motion Congruency: no motion vs. congruent motion vs. incongruent motion) repeated measures ANOVA. There was a significant main effect of Magnitude on time level F (1, 9) = 82.84, MSE = 21406.04, $p < .001$. Not surprisingly, less time was spent level in the 40° condition (M = 3174 ms) than in the 20° condition (M = 3518 ms), simply because trials in the 40° condition began further from level and therefore took longer to correct. There was a significant main effect of motion congruency on time level F (2, 18) = 3.40, *MSE* = 18635.76, $p < .06$, with less time spent level in the incongruent condition. (M = 3285 ms) than in both the no-motion and congruent motion conditions (M = 3356 ms and M = 3396 ms, respectively). There was also a significant Magnitude x Motion Congruency interaction, F (2, 18) = 3.65, *MSE* = 27181.73, $p < .05$. This interaction appears to be primarily driven by the performance enhancement provided by congruent motion cues when visual motion cues are less dramatic (i.e., in the 20° magnitude condition). These data show that congruent motion cueing may be beneficial for attitude recovery in small upset conditions, but less beneficial for recovery from larger attitude upsets.

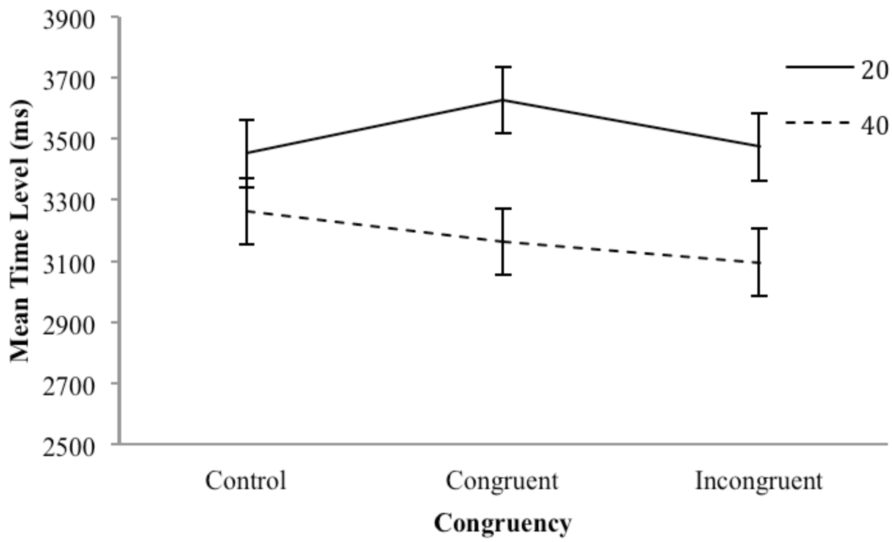


Figure 4. Average mean time level (ms) as a function of motion congruency and magnitude and 95% CI [± 109.5ms].

RMSE. RMSE was calculated as the root of the average squared error (distance from level in degrees) recorded at 60 Hz. RMSE (Figure 5) was analyzed using a 2 (Magnitude: 20° vs. 40°) x 3 (Motion Congruency: no motion vs. congruent motion vs. incongruent motion) repeated measures ANOVA. There was a significant main effect of magnitude F (1, 9) = 189.83, *MSE* = .97, $p < .001$, with more error in the 40° condition (M = 7.17°) than in the 20° condition (M =3.66°). There were no other significant effects.

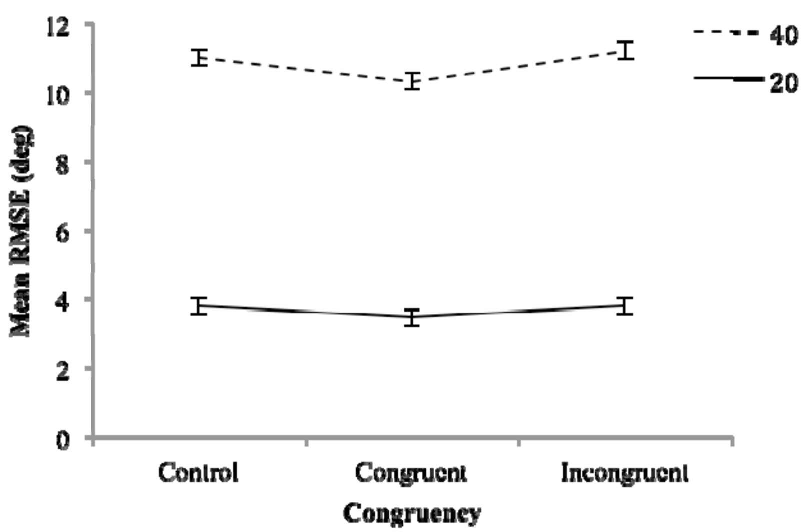


Figure 5. RMSE (deg) as a function of motion congruency and magnitude and 95% CI [± .24 deg].

Conclusion

This study contributes to the research literature on the use of motion cueing in flight training devices. Motion cueing was provided by a motion cueing seat. Incongruent motion cueing was found to result in performance impairments on a basic attitude recovery task as indexed by increased number and time spent in a control reversal, over corrections past level, and decreased time level. It is clear, therefore, that the present experimental paradigm was sensitive and able to index a role for motion cueing on attitude recovery performance. The effect of congruent motion on attitude recovery, however, was not significant in the present research. There are several possible reasons why congruent motion cueing did not impact performance in the present study: (1) the use of congruent motion cues may have been dampened because no-motion, congruent and incongruent motion trials were randomly presented, (2) an outside-in attitude indicator was utilized, which although intuitive, is less representative of commonly used attitude indicators, (3) the onset of the attitude displacements were predicable whereas in flight simulator training unusual attitudes are often introduced without specific countdowns, (4) the attitude recovery task was simplified to left/right bank and did not include pitch or yaw displacements, and (5) the overall task workload was minimal insofar as other concurrent flight-tasks were not present.

Acknowledgments

This research was supported through funding from the Canadian Foundation for Innovation (CFI) to Dr. C. M. Herdman. Additional infrastructure for this research was provided by the Canadian Department of National Defence. Special thanks to Dr. M. Brown for his assistance.

References

Burki-Cohen, J., Sparko, A., & Bellman, M. (2011). Flight simulator motion literature pertinent to airline-pilot recurrent training and evaluation. *AIAA Modeling and Simulation Technologies Conference*, 8-11 August, 2011.

de Winter, J., Dodou, D., & Mulder, M. (2012). Training effectiveness of whole body flight simulator motion: A comprehensive meta-analysis. *The International Journal of Aviation Psychology, 22*(2), 164-183.

Grant, P., Yam, B., Hosman, R., & Schroeder, J. (2006). Effect of simulator motion on pilot behavior and perception. *Journal of Aircraft, 43*(6), 1914-1924.

Hays, R., Jacobs, J., Prince, C., & Salas, E. (1992). Flight simulator training effectiveness: A meta-analysis. *Military Psychology, 4*(2), 63-74.

Holt, L., Schreiber, B., Duran, J., & Schroeder, M. (2011). Evaluating the impact of dynamic fidelity on performance. *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*, 28-1 November-December, 2011.

Jarmasz, J., & Hollands, J. G. (2009). Confidence intervals in repeated-measures designs: The number of observations principal. *Canadian Journal of Experimental Psychology,* 63, 124-137. doi: 10.1037/a0014164

Jian, G., Shutao, Z., & Hongren, L. (2011). Assessment of flight simulator fidelity for pitch task. *Key Engineering Materials, 460-461*, 569-573.

Meyer, G., Wong, L., Timson, E., Perfect, P., & White, M. (2012). Objective fidelity evaluation in multisensory virtual environments: Auditory cue fidelity in flight simulation. *Public Library of Science, 7*(9), e44381.

Pasma, D., Grant, S., Gamble, M., Kruk, R., & Herdman, C. (2011). Utility of motion and motion-cueing to support simulated in-flight rotary-wing emergency training. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55*(1), 133-137.

Sutton, D., Skelton, M., & Holt, L. (2010). Application and implementation of dynamic motion seats. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, 29-2 November-December, 2010.

HUMAN-AUTOMATION PERFORMANCE UNDER TIME PRESSURE HAS LIMITED BENEFITS

Casey Tunstall
Stephen Rice
New Mexico State University
Las Cruces, New Mexico

Previous research by Rice and Keller (2009) has supported that time pressure can positively affect operator compliance with an automated device if the automation is highly reliable despite the impact of speed-accuracy trade-off. When given sufficient time, operators aided by highly reliable automation tended to ignore the aid's recommendation and produced human-automation performance levels less than that of the automation itself. When the operators were put under time pressure, they tended to comply with the automation with a performance that equaled or exceeded that of the automation itself. The current study suggests that the time pressure heuristic is only beneficial when the target is difficult to find by manipulating the time pressure and the difficulty of the search task. The results find that the time pressure heuristic is not as beneficial for easy to find targets as it is for difficult to find targets.

Automation is a helpful tool often used to increase human performance. Wickens and Hollands (2000) denote automation as an entity that has the potential to enhance or replace human performance. It has been found that human performance is lower than the combination of human and automation (human-automation); furthermore, the performance of the automation alone is typically stronger than human-automation (e.g., Dixon, Wickens, & McCarley, 2007; Dixon & Wickens, 2006; Rice, 2009; Rice, Trafimow, Clayton & Hunt, 2008; Weigmann, McCarley, Kramer, & Wickens, 2006). In a more specific case of target detection tasks, human performance was higher when paired with a diagnostic automation than without the aid. Although the performance has been found to increase when the operator is paired with automation, it has also been found to decrease human-automation performance may occur (Breznitz, 1984). This could be attributed to the operator ignoring, second-guessing, or relying too heavily on the automation (Breznitz, 1984).

It is important to look into why human-automation performance is lower than that of automation. One factor that has been found to affect the level of reliance on automation is trust. In many cases, trust in automation is often based on a subjective evaluation on how well the automation will help the operator reach their objective (Lee & See, 2004). However, there are several aspects that effects trust of automation, such as the reliability of the automation itself (Weigmann, 2002). Reliance on automation can be detrimental to performance because the operator may often over-rely on the automation even when it is wrong (Parasuraman, Molloy and Singh, 1993). This can be due to the operator becoming overly content with the automation and disregarding errors. The opposite has also been known to occur if the reliance is low the operator may not use the aid at all (Dixon & Wickens, 2006).

It was found that the different errors in automation cause change in the operator's trust of the automation (Meyer, 2001; 2004; Rice, 2009; Parasuraman & Riley, 1997). These errors are known as false alarms and misses. If the automation is more apt to produce false alarms then the compliance decreases. If the automation is more apt to produce misses, the reliance on the automation decreases (Meyer, 2001; 2004). In some cases the operators may be in situations that include multiple tasks, stress, or penalties for not complying. If these situations exist, the operator may be more willing to depend on the automation. Another situation that could influence the operator's decision-making process is the existence of time pressure (Maule, Hockey & Bdzola, 2000; Maule & Svenson, 1993; Maule, 1997; Maule & Edland, 1997). Further research has shown that an operator under time pressure has less severe judgments that generally lead to better decisions (Kaplan, Wanshula, & Zanna, 1993; De Neys, 2006). By incorporating time pressure, the operator may be able to take into account more essential information that is incorporated into their decision making process (Ben-zur & Breznitz, 1981; Bockenholt & Kroeger, 1993; Kerstholt, 1995; Payne, Bettman, & Johnson, 1988; Wright, 1974).

A study by Rice & Keller (2009) participants was asked to search through aerial photographs for an enemy target with the use of an automated aid. The aid that was used varied between reliability percentages. The participants were assigned to one of two conditions. One condition was set up to create the effect of time pressure by only having 2 seconds with the photograph. The other condition allowed the participants 8 seconds in order to avoid time pressure effects. The results showed that when the participants were under time pressure, their reliance on the automated aids were higher than the condition of no time pressure. Also, the results were able to demonstrate that the time pressure condition relied on the automated aid regardless of the reliability percentage. Trust did not vary between the two conditions. The performance of the time pressure condition was higher than the other condition when the reliability of the aid was high but was lower when the reliability percentage was low.

In the current study, we will expand upon the original study by Rice & Keller (2009) by manipulating the difficulty of the aerial photographs. We would like to observe the differences of performance between the difficult and easier task when under time pressure. We hypothesized that when the operator is faced with an easy task, the time pressure effects will not be as strong or beneficial as it would be for a difficult task.

Method

Experiment 1

One hundred and sixty (100 females) undergraduate students at New Mexico State University participated in the experiment for partial course credit. The mean age was 20.51 (SD = 2.59). All participants reported normal or corrected to normal vision. The design of Experiment 1 was a recreation of the previous study by Rice & Keller (2009). A 2 x 5 between-participant study was used with time-pressure and automation reliability as the two independent variables. The time-pressure was either speeded or unspeeded. The automation reliability was presented as 100%, 95%, 80%, 65%, or no automation "B". Each participant was randomly selected to one of 10 conditions.

Each participant was presented with 100 photographs of Baghdad (see Figure 1). Fifty of the photographs contained an enemy tank, and the other 50 photos did not. The experiment was conducted on a Dell computer with a 20" monitor using 1024 x 768 resolution via E-Prime 1.1. The participants were seated 21" from the display by using a chin rest to control the position. They were asked to read through the instructions on E-Prime and to press any key once they were ready to begin. The participants were commanded to press the "J" key if they detected a tank. If they determined no tank was present, they were commanded to press the "F" key. The participants were instructed to try to maintain the highest accuracy as possible. They were informed if an automated aid were present to relay recommendations on whether or not an enemy tank was present in the photograph. Depending on which condition the participant was in; they were told the reliability percentage of the aid (100%, 95%, 80%, 65%, no automation) and how much time they had to view the image (2 seconds or 8 seconds) after the recommendation had been given by the automated aid.

Once the experiment began, a slide provided the recommendation of the automated aid for 1000 ms prior to the aerial photograph slide. The recommendation was one of two different messages, "The automation has detected a tank!" or "The automation has determined that there is no tank present!" The aerial photograph was shown for 2 seconds for the speeded condition and 8 seconds for the unspeeded condition. The participants were presented with instant feedback to each response for 1000 ms after each key press. The feedback included whether they were correct (in green letters) or incorrect (in red letters) and current percent of correct answer responses.

Figure 1. A sample image used in the experiment. The tank is located in the top central area to the southeast of the concentric circles.

Results

Figure 2 presents the data from the experiment. For the purposes of data analysis, all accuracy scores were converted to d' measures (signal detection theory). There was a main effect of Condition, $F(4, 150) = 108.18$, $p < .001$, but no main effect of Speeded-Unspeeded (SU), $F(1, 150) = 1.76$, $p = .19$. An interaction between Condition and SU, $F(4, 150) = 5.74$, $p < .001$, revealed that the benefits of the speeded heuristic were limited to the higher reliability levels (100% and 95%). These results replicate those found by Rice and Keller (2009).

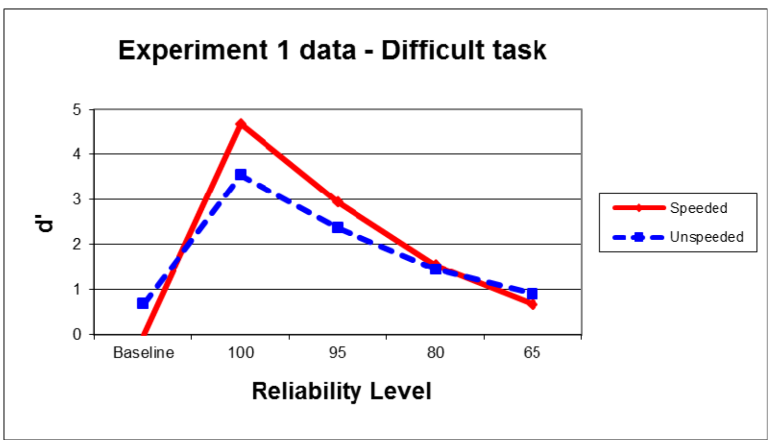


Figure 2. Data from Experiment 1.

Experiment 2

Method

One hundred and forty (100 females) undergraduate students at New Mexico State University participated in the experiment for partial course credit. The mean age was 20.00 (SD = 2.63). All participants reported normal or corrected to normal vision. One participant's data was dropped from analysis due to a corrupted file.

Experiment 2 was identical to Experiment 1 with one exception. Instead of using the images from Experiment 1, new images that were less cluttered were used. A pilot study showed that the targets in these images were easier to find than those used in Experiment 1.

Results

Figure 3 presents the data from Experiment 2. There was a main effect of Condition, $F(4, 129) = 15.70$, $p < .001$, and a main effect of Speeded-Unspeeded (SU), $F(1, 129) = 38.88$, $p < .001$. The interaction between Condition and SU, $F(4, 129) = 1.83$, $p = .13$, was not significant. These data reveal that when the task of finding the tank is easy, then there is no benefit to using the speeded heuristic at any level of reliability.

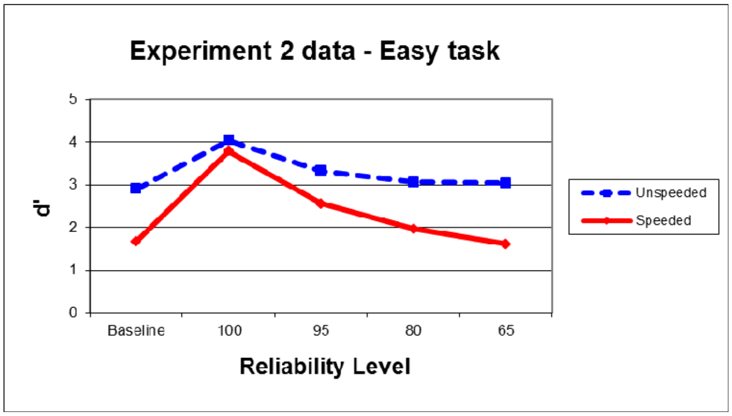


Figure 2. Data from Experiment 2.

Discussion

The purpose of the current study was to extend the findings from Rice and Keller (2009) and Rice and Trafimow (2012), by examining whether the beneficial effects of the time pressure heuristic would remain despite decreasing the difficulty of the task in question. We manipulated both the amount of time participants were allowed to view the aerial images and the difficulty of the search task. We hypothesized that when the task was difficult, time pressure would result in increased human-automation performance, replicating previous findings, but that when the task was easy, the time pressure heuristic would not be as beneficial to performance because participants would perform just as well on their own. The findings from the study showed that our hypotheses were confirmed. When the search task was difficult, the data replicate those from Rice and Keller (2009). However, when the task was easy, the time pressure did not benefit performance, but in fact, decreased performance compared to when there was no time pressure.

The differences between the easy task and the hard task performances could be due to a few explanations. The time pressure heuristic often drives the operator to use the automation in order to make a decision (Rice and Keller, 2009). If the task is easy, the time pressure heuristic may not influence the operator in their decision or use of the automation. In other words, the operator may not feel pressure due to the small amount of time with the photographs and can confidently make a decision without employing the automation. A second explanation could be that the operator is over confident in their decision due to the task being easy. If the operator assumed that they are correct every time, they are less likely to agree with the automation, which then drives down their performance in cases where the automation's reliability is high.

A study by Rice and Trafimow (2012) also expanded upon the Rice and Keller (2009) study by applying the potential performance theory (Trafimow & Rice, 2008, 2009). They were able to analyze the differences between the speeded group and unspeeded group. The observed scores, potential scores, and the consistency scores were all used in order to perform the PPT analyses (Rice and Trafimow, 2012). They were able to show that the time pressure task allowed for a better performance due to the decrease in randomness when making a decision. This could help explain why the easier task had a lower performance. The randomness in the task increases when dependence on the automation decreases. In other words, if the operator had used the automation to make their decision, the randomness would have been lower and the performance higher.

Practical implications can be taken from these newfound results. Since it is known that the automation has the potential to increase performance, it is important to know what factors can influence the operator's reliance and compliance with the automation. If a time pressure heuristic is put into place in the design of such automation and tasks, it is important to know that it will not always increase the performance of the operator if the task is too simple. This should be kept in mind when training and designing automation aids.

Conclusion

The results of the first experiment concurred with the original study by Rice and Keller (2009). Imposing a time pressure heuristic to a task was found to increase performance. The second experiment found that lessening the difficulty of the task decreases the effectiveness of the time pressure heuristic when the automation reliability is high. The easier task decreased the operator's dependence on the automation which in turn decreased performance.

Acknowledgements

We would like to acknowledge all of the New Mexico State University undergraduates that assisted in the lab and collected the data.

References

Ben-Zur, H., & Breznitz, S. J. (1981). The effect of time pressure on risky choice behavior. *Acta Psychologica*, *47*(2), 89-104.

Bockenholt, U., & Kroeger, K. (1993). The effect of time pressure in multiattribute binary choice tasks. In O. Svenson & A.J. Maule (Eds.), *Time pressure and stress in human judgement and decision making* (pp. 195-214). New York: Plenum Press

Breznitz, S. (1984). *Cry wolf: The psychology of false alarms* (pp. 1-100). Hillsdale, NJ: Lawrence Erlbaum Associates.

De Neys, W. (2006). Dual Processing in Reasoning Two Systems but One Reasoner. *Psychological Science*, *17*(5), 428-433.

Dixon, S. R., & Wickens, C. D. (2006). Automation Reliability in Unmanned Aerial Vehicle Control: A Reliance-Compliance Model of Automation Dependence in High Workload. *Human Factors*. *48*(3), 474-486.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors*. *49*(4), 564-572.

Kaplan MF, Wanshula LT, Zanna MP. 1993. Time pressure and information integration in social judgment: the effect of need for structure. In *Time Pressure and Stress in Human Judgement and Decision Making*, ed. O Svenson, J Maule, pp. 255–67. New York: Plenum

Kerstholt, J. H. (1995). Decision making in a dynamic situation: the effect of false alarms and time pressure. *Journal of Behavioral Decision Making*, *8*(3), 181-200.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46,* 50-80.

Maule, A. J. (1997). Strategies for adapting to time pressure. *Decision making under stress- Emerging themes and applications (A 99-12526 01-53), Aldershot, United Kingdom, Ashgate, 1997*,, 271-279.

Maule, A.J., & Edland, A.C. (1997). The effects of time pressure on judgement and decision making. In R. Ranyard, W. R. Crozier & O. Svenson, *Decision making: cognitive models and explanation* (pp. 189-204). London: Routledge & Kegan Paul.

Maule, A. J., Hockey, G. R. J., & Bdzola, L. (2000). Effects of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy. *Acta Psychologica*, *104*(3), 283-301.

Maule, A. J., & Svenson, O. (Eds.). (1993). *Time pressure and stress in human judgment and decision making*. Springer.

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *43*(4), 563-572.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(2), 196-204

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency". *International Journal of Aviation Psychology, 3*(1), 1-23.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39,* 230-253.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534.

Rice, S. (2009). Examining single-and multiple-process theories of trust in automation. *The Journal of general psychology*, *136*(3), 303-322.

Rice, S., & Keller, D. (2009). Automation reliance under time pressure. *International Journal of Cognitive Technology*, *14*(1), 36.

Rice, S., & Trafimow, D. (2012). Time pressure heuristics can improve performance due to increased consistency. *The Journal of General Psychology*, 139(4), 273-288.

Rice, S., Trafimow, D., Clayton, K., & Hunt, G. (2008). Impact of the contrast effect on trust ratings and behavior with automated systems. *Cognitive Technology Journal*, *13*(2), 30-41.

Wiegmann, D. A. (2002). Agreeing with automated diagnostic aids: A study of users' concurrence strategies. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*(1), 44-50.

Wiegmann, D., McCarley, J. S., Kramer, A. F., & Wickens, C. D. (2006). Age and automation interact to influence performance of a simulated luggage screening task. *Aviation, space, and environmental medicine*, *77*(8), 825-831.

Wickens, C. D., & Hollands, J. G. Engineering Psychology and Human Performance. 2000. *ISBN: 0-321-04711-7*.

Wright, P. (1974). The harassed decision maker: Time pressures, distractions, and the use of evidence. *Journal of applied psychology*, *59*(5), 555.

APPLYING THE COMPLIANCE-RELIANCE MODEL TO SYSTEM-WIDE TRUST THEORY IN AN AVIATION TASK

Casey Tunstall
Kasha Geels-Blair
Stephen Rice
New Mexico State University
Las Cruces, New Mexico

Diagnostic automation is frequently used to assist pilots. Despite extensive research on trust in automation, multiple automation research is sparse. Two recent studies focused on effects of trust in multiple automation using system-wide trust (SWT) theory (Keller & Rice, 2010; Rice & Geels, 2010). According to SWT, operators treat multiple aids as one entity, rather than separate entities; an unreliable aid negatively affects trust in nearby aids. We combined SWT with Meyer's (2001, 2004) compliance-reliance model in order to show whether contagion effects of FAs and misses differ. Participants monitored 8 gauges for system failures. Each gauge had an automated aid that provided recommendations. The first aid was 70% reliable, and produced either FAs or misses, while the other 7 aids were 100% reliable. The data indicate that both FAs and misses cause contagion on the perfectly reliable aids, but FAs caused more contagion in both compliance and reliance measures.

It is common for operators to interact with diagnostic aids in the aviation industry. While much research has focused on operator trust in a single aid, very little has analyzed the effects of trust in multiple aids. Keller and Rice (2010) recently proposed a theory called system-wide trust (SWT), whereby operators tend to treat multiple aids as part of a "system" rather than as individual components. When one of the aids is unreliable, it affects trust in the other aids, even when the other aids are reliable and numerous (Keller & Rice, 2010; Rice & Geels, 2010). The purpose of the current study is to analyze the differential effects of automation false alarms (FAs) and misses on SWT. By combining SWT theory with Meyer's (2001, 2004) compliance-reliance model, we hope to show that the contagion effect of FAs and misses on reliable aids are qualitatively different.

System-Wide Trust Theory

Although much research has focused on a single automated aid (Rice, 2009; Maltz & Shinar, 2003; Meyer, 2001, 2004; Dixon & Wickens, 2006; Dixon, Wickens, & McCarley, 2007; Lee & Moray, 1994; Parasuraman, Molloy, & Singh, 1993), the effect that an unreliable aid would have on nearby automated aids was unknown before the introduction of system-wide trust theory (Keller & Rice, 2010; Rice & Geels, 2010). Keller and Rice (2010) proposed two possibilities when pairing an unreliable automated aid with a perfectly reliable aid. The first possibility, component specific trust (CST), states that participants would treat the automated aids differently according to their level of reliability. The second possibility states that the two automated aids would be treated as a "system" in which trust would be calibrated to both automated aids as one entity. Therefore, if one automated aid is unreliable, it may cause a negative contagion effect that spreads to nearby, perfectly reliable, aids.

Keller and Rice (2010) had participants perform a simulated unmanned aerial systems task while also monitoring for failures in two gauges. The automated aids were 70%, 85%, or 100% reliable, only erring with FAs. Participants could disagree with the automated aids if they wished, but the final decision was up to them. Performance was recorded using response times and accuracy. The data supported the second possibility. System-wide trust took place and participants' trust in the perfectly reliable aid dropped to levels of the unreliable aid.

The findings were further tested in a follow-up SWT study (Rice & Geels, 2010) using four aids and a single-task paradigm (to possibly remove any dual-task confounds). Although three of the aids were 100% reliable, the left-most aid was either 70% or 100% reliable, with the unreliable condition erring by producing only misses. In lieu of using accuracy, agreement rates were used along with the continued use of response times in order to show the effects on trust. The findings supported Keller and Rice (2010); even in conditions where participants had explicit knowledge of the reliability of each automated aid, SWT effects still occurred.

Compliance and Reliance

Much research has compared the effects of FAs and misses on operator trust (e.g. Bainbridge, 1983; Lee & Moray, 1994; Lee & See, 2004; Meyer, 2001, 2004; Parasuraman & Riley, 1997; Parasuraman, Sheridan, & Wickens, 2000). For the current study, we look to Meyer's (2001, 2004) model of compliance and reliance in order to test for differential effects of the two errors on trust in multiple automation. This model predicts that FAs will have a negative effect on compliance, which is defined as behaviors toward alerts, while misses will have a negative effect on reliance, defined as behaviors toward non-alerts. The model predicts that errors will only negatively affect their respective behaviors: false alarms will only affect compliance and misses will only affect reliance.

Three studies (Dixon & Wickens, 2006; Dixon, Wickens, & McCarley, 2007; Rice, 2009) were performed in order to test the compliance-reliance model. The first study (Dixon & Wickens, 2006) had participants perform a simulated unmanned aerial systems mission while comparing the effect of FAs and misses in a systems monitoring task. FAs not only reduced compliance, but also negatively affected reliance. Misses only had a negative effect on reliance. The second study (Dixon, Wickens, & McCarley, 2007) tested the findings using a more controlled laboratory setting. Dixon, Wickens, & McCarley (2007) pointed out that, in these first two studies, the presentation of FAs and misses were not equivalent. FAs were immediately noticed because an alert sounded when the automation detected a failure, while a miss was silent because non-alerts were periods of silence from the automation. This could explain the differential effects between FAs and misses in the first two studies.

In the third study, Rice (2009) alleviated this effect by making FAs and misses equally salient. The dual-task confound was also removed by having participants perform a single visual search task. Participants searched for a helicopter with the help of an automated aid that presented a recommendation before each trial. The automated aid was either FA-prone or miss-prone and varied from 55% to 100% reliable in 5% increments. Both types of errors affected compliance and reliance, with FAs strongly affecting compliance and misses strongly affecting reliance. This was the first study to show that misses can have an effect on compliance.

The Current Study: Merging System-Wide Trust Theory with the Compliance-Reliance Model

Although the previous studies examined either the contagion effects of an unreliable aid on nearby reliable aids (Keller & Rice, 2010; Rice & Geels, 2010) or the effects of FAs and misses on a single aid (Dixon & Wickens, 2006; Dixon, Wickens, & McCarley, 2007; Rice, 2009), the effects of FAs and misses on multiple automated aids have not yet been compared. The current study aims to do just that by examining SWT (Keller & Rice, 2010) using the compliance-reliance model (Meyer, 2001, 2004). By combining the two, we can see the effects of FAs and misses on compliance and reliance, whether a contagion effect still takes place, and whether a difference in contagion to nearby aids exists between FAs and misses.

In this experiment, participants were responsible for monitoring 8 gauges for system failures. Each gauge was augmented by an automated aid that provided recommendations. The first aid was either 70% or 100% reliable, and failures were either FAs or misses, while the remaining 7 aids were always 100% reliable. Trust was measured using agreement rates and response times to both alerts and non-alerts. We hypothesize that 1) FAs and misses will affect both compliance and reliance in the first gauge, as described in Rice (2009), 2) that SWT would still occur, despite the increase in perfectly reliable aids from previous SWT studies (Keller & Rice, 2010; Rice & Geels, 2010), resulting in the strongest test of SWT, and 3) that contagion for the aids on gauges 2-8 would be differentially affected by FAs and misses. FAs should have a stronger effect on compliance with a weaker effect on reliance and misses should have the reverse effect.

Method

One hundred thirty-three (84 females) undergraduates participated for partial course credit; mean age was 20.27 (*SD* = 4.75). All participants were tested for normal or corrected-to-normal vision and were assigned to one of four conditions: Baseline (no automation present), 100% (all 8 automated aids were perfectly reliable), 70M (automation for Gauge 1 was 70% reliable and produced only misses, while all other aids were 100% reliable), or 70F (like 70M, except the 70% reliable aid produced only FAs). Due to the absence of automated aids, the Baseline condition was excluded from analyses. Participants were first seated at a computer and given scripted verbal instructions, followed by 5 practice trials. Once comfortable with the task, participants were told whether there would be automated aids present during the experiment, as well as the reliability of each. Viewing distance was controlled for by using a chinrest situated approximately 21” from the computer monitor.

There were 50 experimental trials which lasted 20 seconds each, followed by a feedback screen, which lasted for 5 seconds. The display consisted of eight simulated gauges with independent random values; each augmented by an automated aid (see Figure 1). Gauges were read by discerning the values given by each of three hands. The short, fat hand represented units of 1000, the long, fat hand represented units of 100, and the long, thin hand represented units of 10. Above each gauge was a range indicator which displayed random 4- and 3-digit values that changed for every trial. To determine the “safe range”, a range indicator that read “1500 (500)” would mean that the safe range is 1500 +/- 500. The automated aids gave a recommendation of whether the corresponding gauge for each was considered safe or if it had failed. The aids were located underneath the gauges and displayed a green box to state that the gauge was safe or a red box to state that it was a failure. Although the automation gave a recommendation, participants were told the final decision rested with them. They made their decision for each gauge by clicking on either the “Safe” or “Failure” options, which were located below each gauge. Once they clicked on an option for all of the gauges, a “Continue” button appeared so that participants could continue to the feedback screen. Feedback was given by outlining the choice display for each gauge. A green outline meant the participant chose correctly and a red outline meant they answered incorrectly. The experiment lasted approximately 40 minutes. Upon completion, participants were debriefed and dismissed.

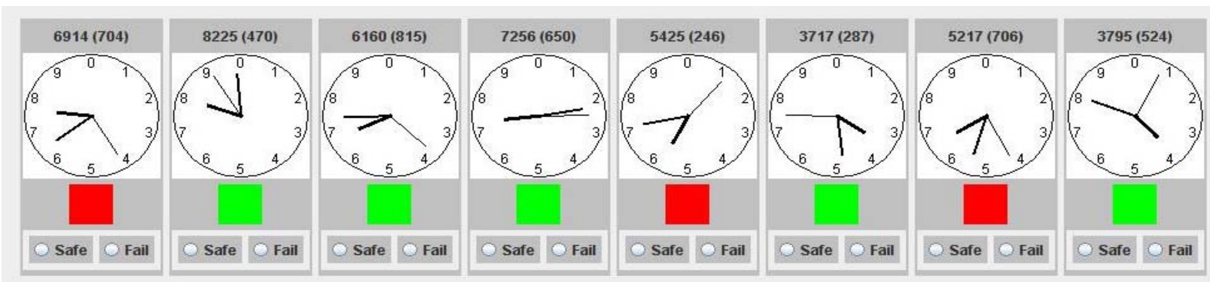


Figure 1. Gauge display. The display was originally presented in color.

Results

Gauge 1 Performance

Agreement Rates. Agreement rates were measured by the percentage in which participants agreed with the automated aid. We split up the agreement rates for alerts and non-alerts, as can be seen in Figure 2. A 2-way ANOVA using Condition and Aid showed a main effect of Condition, $F(2, 111) = 72.97$, $p < .001$, $\eta^2 = .57$, which shows that performance in the 100% condition was superior to the unreliable conditions, $t(111) = 11.99$, $p < .001$. The interaction between Condition and Aid was also significant, $F(2, 111) = 10.10$, $p < .001$, $\eta^2 = .15$. After removing the 100% condition data, we looked at the interaction between the 70F and 70M conditions and the Alert-NonAlert factor, $F(1, 74) = 13.51$, $p < .001$, $\eta^2 = .15$. This showed that for alerts, compliance suffered more in the 70F condition as compared to the 100% condition, $p < .001$, but it also suffered significantly in the 70M condition, $p < .001$. For NonAlerts, reliance suffered most in the 70M condition compared to the 100% condition, $p < .001$, but it also suffered significantly in the 70F condition, $p < .001$. This crossover effect supports Hypothesis 1.

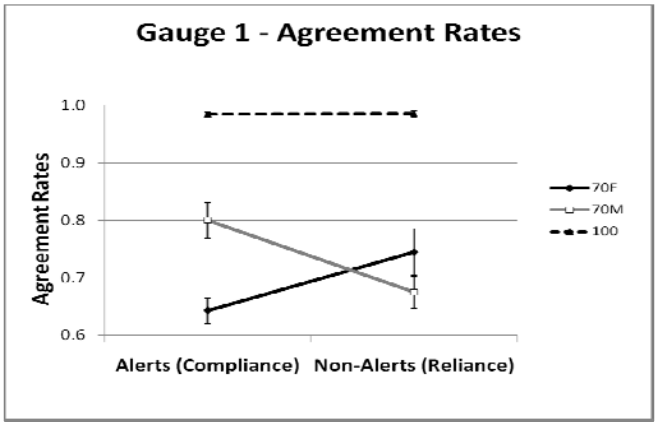


Figure 2. Agreement rates for compliance and reliance for all conditions. SE bars are included.

Response Times. Response times were measured for each gauge since participants responded one at a time. RT was collected starting from the beginning of the experiment until the participant responded to any gauge. After each response, the clock was reset so that RTs only reflected the time between responding to each gauge. A 2-way ANOVA using Condition and Alert-NonAlert as factors showed a main effect of Condition, $F(2, 111) = 14.11$, $p < .001$, $\eta^2 = .20$, which means that responses in the 100% condition was superior to the unreliable conditions, $t(111) = 5.28$, $p < .001$. A main effect of Alert-NonAlert was also present, $F(1, 111) = 6.18$, $p < .05$, $\eta^2 = .05$, which indicates faster RTs to NonAlerts than Alerts. There was also a significant interaction between the two factors, $F(2, 111) = 5.77$, $p < .01$, $\eta^2 = .09$. It should be noted that *t*-tests conducted between the 70M and 70F conditions revealed no differences between the conditions for both alerts and non-alerts (both *p*s > .10).

Contagion Effects

Agreement Rates: Alerts. Compliance was measured using the agreement rate to alerts (Figure 3). Using Condition and Aid as factors, a 2-way ANOVA indicated a main effect of Condition, $F(2, 111) = 10.53$, $p < .001$, $\eta^2 = .16$, a main effect of Aid, $F(7, 777) = 59.17$, $p < .001$, $\eta^2 = .35$, and a significant interaction between the two, $F(14, 777) = 17.12$, $p < .001$, $\eta^2 = .24$. The drop in performance as compared to the 100% condition was not equal across all Aids. Further analysis was done on Aids 2-8 in order to further test Hypotheses 2 and 3. A 2-way ANOVA using Condition and Aids 2-8 as factors showed a main effect of Condition, $F(2, 111) = 3.27$, $p < .05$, $\eta^2 = .06$, this indicated a drop in agreement rates in the 70F condition, $F(1, 74) = 4.29$, $p < .05$, $\eta^2 = .06$, and the 70M condition, $F(1, 74) = 4.06$, $p < .05$, $\eta^2 = .05$, as compared to the 100% condition. The Interaction between Condition and Aid was marginally significant, $F(12, 666) = 1.76$, $p = .051$, $\eta^2 = .03$. Since response times mirrored agreements rates and due to lack of space, the RT analyses will not be reported here.

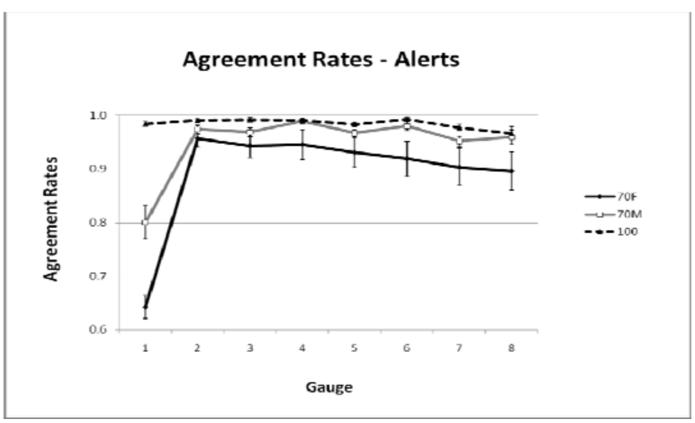


Figure 3. Compliance. SE bars are included.

Agreement Rates: Non-Alerts. Reliance was measured by how often the participant agreed with the automated aid for non-alerts (Figure 4). A 2-way ANOVA with Condition and Aid as factors showed a main effect of condition, $F(2, 111) = 8.02$, $p = .001$, $\eta^2 = .13$, a main effect of Aid, $F(7, 777) = 53.68$, $p < .001$, $\eta^2 = .33$, and a significant interaction between Condition and Aid, $F(14, 777) = 14.55$, $p < .001$, $\eta^2 = .21$, which means a drop in performance from 100% condition to the unreliable conditions was not equal across all Aids. Further analysis was done on the reliable Aids in order to test hypotheses 2 and 3. A 2-way ANOVA using Condition and Aids 2-8 as factors showed a marginally significant main effect of Condition, $F(2, 111) = 2.65$, $p = .075$, $\eta^2 = .05$,which indicates a drop in performance for these Aids in the 70F condition, $F(1, 74) = 3.82$, $p = .054$, $\eta^2 = .05$, and the 70M condition, $F(1, 74) = 4.47$, $p < .05$, $\eta^2 = .06$, compared to the 100% condition. There was a significant main effect of Aid, $F(6, 666) = 6.71$, $p < .001$, $\eta^2 = .06$.

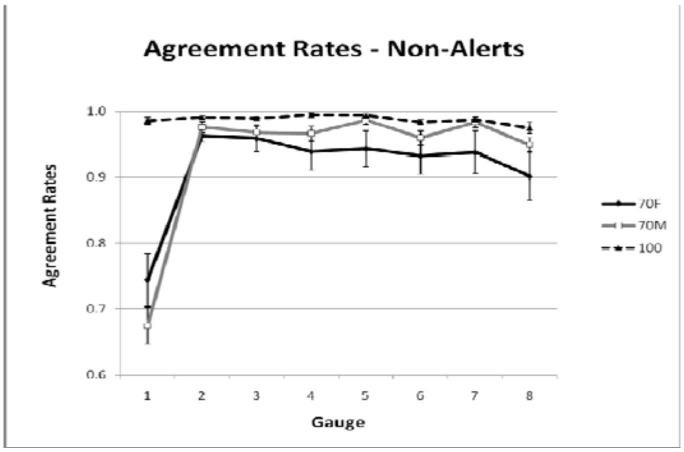


Figure 4. Reliance. SE bars are included.

Discussion

The purpose of the current study was to replicate the findings of Rice (2009) in order to test the differential effects of FAs and misses on compliance and reliance (Meyer, 2001, 2004), to find whether a contagion effect would still take place (Keller & Rice, 2010; Rice & Geels, 2010), despite the increase in reliable automated aids, and to discover the effects of FAs and misses in multiple automation. We tested three hypotheses based on findings in the previous literature. Our first prediction was that FAs would have a stronger negative effect on compliance, while also affecting reliance, and misses would have the reverse effect on the unreliable aid (Rice, 2009). Second, we predicted that SWT would still occur, despite the increase in reliable aids from previous SWT studies (Keller & Rice, 2010; Rice & Geels, 2010). Lastly, we predicted that FAs and misses in the first aid would differentially affect compliance and reliance throughout all the automated aids, with FAs having a greater negative effect on compliance and a weaker effect on reliance, while misses would have the reverse effect.

Hypotheses 1 and 2 were supported by the data, while Hypothesis 3 was only partially supported. With regards to Hypothesis 1, both error types had effects on compliance and reliance in the first aid, as seen in Rice (2009). The previous study had been the first to discover that both errors could reduce both types of trust and we have replicated the findings here. Pertaining to Hypothesis 2, SWT still took place, despite the increase in automated aids which has been shown to take place with 2, 4, and now 8 automated aids, though not as strong as we had predicted (Keller & Rice, 2010; Rice & Geels, 2010). The data indicate that both FAs and misses cause considerable contagion on the perfectly reliable aids, as predicted by SWT theory. Unfortunately, in regards to Hypothesis 3, FAs appear to cause more contagion than misses in both compliance and reliance measures, while misses did not have such a strong effect. This finding does not reflect the effect on trust in the first automated aid.

Conclusion

SWT theory (Keller & Rice, 2010) and Meyer's (2001, 2004) compliance-reliance model were combined in order to assess trust in multiple automated aids. By combining both theories, we were able to discover that FAs and misses replicate the mirrored effect on compliance and reliance (as seen in Rice, 2009) and that FAs may have caused a stronger contagion effect across all the perfect aids as compared to misses. Designers should be concerned with the possibility that operators may group multiple automated aids as an individual system rather than as independent components. Furthermore, designers should also be aware that the contagion effects may be stronger on reliable aids adjacent to an automated aid producing FAs than one producing misses.

Acknowledgments

The authors wish to thank Joshua Sandry and Gayle Hunt for their helpful comments, Sandra Deming and Laurel Ashley for their help in collecting data, and Chandrasekhar Thotakura for programming the simulator.

References

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19,* 775-779.

Dixon, S. R., & Wickens, C. D. (2006). Automation Reliability in Unmanned Aerial Vehicle Control: A Reliance-Compliance Model of Automation Dependence in High Workload. *Human Factors*. *48*(3), 474-486.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors*. *49*(4), 564-572.

Keller, D. & Rice, S. (2010). System-wide versus component-specific trust using multiple aids. *Journal of General Psychology*, *137*(1), 114–128.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40,* 153-184.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46,* 50-80.

Maltz, M., & Shinar, D. (2003). New alternative methods in 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*(4), 563-572.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors, 46*(2), 196-204.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency". *International Journal of Aviation Psychology, 3*(1), 1-23.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39,* 230-253.

Parasuraman, R., Sheridan, T., & Wickens, 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.

Rice, S. (2009). Examining single and multiple-process theories of trust in automation. *Journal of General Psychology, 136*(3), 303-319.

Rice, S. & Geels, K. (2010). Using system-wide trust theory to make predictions about dependence on four diagnostic aids. *Journal of General Psychology, 137*(4), 362-375.

CLARIFYING COGNITIVE COMPLEXITY AND CONTROLLER STRATEGIES IN DISTURBED INBOUND PEAK ATC OPERATIONS

Marian J. Schuver-van Blanken
Mariska I. Roerdink
Air Traffic Control the Netherlands (LVNL), Human Factor Department
Schiphol airport, Amsterdam, the Netherlands

Air traffic controller (ATCo) expertise is crucial in safely and effectively managing operational disturbances and unpredictable events. The high level of ATCo expertise needed in these situations originates from the cognitive complexity in the ATC task. To cope with cognitive complexity in managing operational disturbances, controllers apply strategies to avoid task performance being compromised. Using the ATCo Cognitive Process and Operational Situation (ACoPOS) model, this paper clarifies the cognitive complexity involved in disturbed inbound peak operation within dense airspace for Schiphol airport at ATC the Netherlands (LVNL). Complexity issues in cognitive processes and operational factors involved are described. Strategies used by expert controllers in response to day to day disturbances in inbound peak operation are described based on results of a focus group. Results indicate the existence of new strategies, supplementary to those found in literature.

Within a wide range of operational situations, air traffic controllers are able to ensure safety whilst keeping the additional goals of efficiency and environment in optimal balance. Especially in managing operational disturbances and unpredictable events, air traffic controller (ATCo) expertise is crucial (Redding, Ryder, Seamster, Purcell & Cannon, 1992; Schuver-van Blanken, Huisman & Roerdink, 2010; SESAR, 2007). However, acquiring the level of expertise needed for handling these situations is one of the main drop-out reasons in air traffic control (ATC) training (Oprins, 2008). The high level of ATCo expertise originates from the cognitive complexity that is inherent in disturbed ATC operation.

Day to day air traffic control is frequently characterized by disturbed operation as a result of dynamic factors in the situation, unpredictable events or complex situations, not necessarily being extreme or exceptional situations. In disturbed operation, traffic handling has to be (temporarily) adjusted or traffic streams have to be (temporarily) rebuild to mitigate the disturbance, while safety is ensured and optimal efficiency is aimed for. The resulting cognitive complexity is determined by both the complex cognitive processes involved as well as the characteristics of the operational situation. To cope with cognitive complexity in disturbed operation, controllers continuously use strategies to adapt their task performance in response to the characteristics and dynamics of the operational situation. Despite the fact that the importance of controller strategies in ATC performance is emphasized in literature, studies that deal with these issues are limited (Fothergill & Neal, 2008; Nunes & Mogford, 2003; Schuver-van Blanken & van Merriënboer, 2012). This paper provides insight in cognitive complexity involved in a situation that is considered prototypical for handling daily disturbances and unpredictable events at ATC the Netherlands (LVNL): inbound peak operation in dense airspace for Schiphol airport. Next, the paper describes the strategies expert controllers apply in response to disturbed inbound peak operation to mitigate cognitive complexity.

The ATCo Cognitive Process & Operational Situation model (ACoPOS)

To analyse and clarify cognitive complexity in ATC, the ATCo Cognitive Process & Operational Situation (ACoPOS) model was developed at ATC the Netherlands (LVNL) (see Figure 1) (Schuver-van Blanken, Huisman & Roerdink, 2010). The ACoPOS model extends the competences of LVNL's ATC performance model (Oprins, Burggraaff & van Weerdenburg, 2006) with elements in the operational ATC situation. This way, cognitive complexity issues can be pinpointed as cognitive processes cannot be seen separately from the context and operational situation in which the tasks are performed. The ACoPOS model was developed based on a literature review in ATC complexity (e.g. Mogford, Guttman, Morrow & Kopardekar, 1995; Hilburn, 2004) as well as the models of Endsley (1995) and Histon and Hansman (2008), together with practical operational experiences. Distinguished in the ACoPOS model are cognitive processes (right-hand side of the model) and the operational situation (left-hand side of the model). By means of the ACoPOS model a picture can be drawn of the ATCo cognitive processes in a certain operational situation and the factors causing cognitive complexity. The model will be explained in the following sections by means of a description of prototypical inbound peak operation in the Amsterdam ACC South sector.

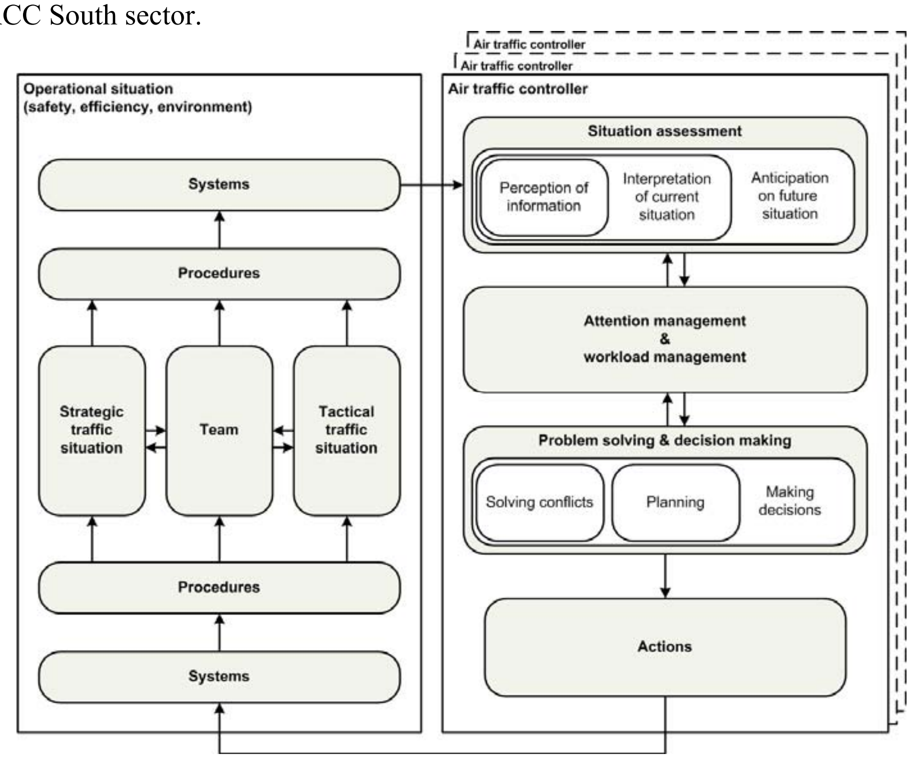


Figure 1, The ATCo Cognitive Process & Operational Situation model (ACoPOS model) (adapted from Schuver-van Blanken, Huisman & Roerdink, 2010)

Operational situation: Inbound peak operation in the Amsterdam ACC South sector

In the current operation at Schiphol Airport, Amsterdam, inbound traffic is delivered at the initial approach fixes (IAFs) by the Amsterdam Area Control Centre (AMS ACC). From the IAFs, tactical vectoring is applied by the Schiphol Approach controllers (SPL APP) to guide traffic to one of the runways. Five area control sectors exist that feed three IAFs. The three IAFs are assigned to a landing runway, with a maximum of two landing runways available at the same time, therefore multiple IAFs merging to a single runway. The South sector is the only sector that feeds traffic to the RIVER IAF. RIVER is different from the other two IAFs in that traffic from this IAF is usually used to balance the traffic amount over both runways, meaning aircraft flying inbound from this sector will often land on different runways and are merged with traffic from the other IAFs.

The ACoPOS model includes factors that define the situation and constitute ATC complexity. The model can be used to provide a structured overview of a prototypical inbound peak situation in the AMS South sector described above, including factors relating to complexity. This overview and these factors are typically determined at ATC the Netherlands by combining human factors analysis with consulting (expert) air traffic controllers. The description below provides an overview of prototypical inbound peak operation in the morning, characterised by many operational disturbances and unpredictable events that have to be handled and part of daily operational practice.

- *Strategic traffic situation:* The strategic traffic situation sets the framework within which traffic has to be handled, such as the traffic volume, the airspace and the runways in use. For inbound peak situation a maximum of 2 landing- and 1 take-off runway is used. Different configurations of runway-use in the inbound peak exist. Further, the South sector has a very limited airspace and traffic volumes in peak operations can be high. Restrictions on declared capacity may apply.
- *Tactical traffic situation:* The tactical traffic situation is characterised by the dynamic nature of the actual situation. Typically, the morning inbound peak is characterised by bunches of aircraft arriving at the same time. Variations in traffic, e.g. crossing and regional traffic or slow climbing traffic (aircraft performance), impact traffic flows. Standard arrival routes are mainly used but interactions with outbound and crossing routes exist. In addition, the wind- and weather-situation impacts traffic handling and can result in traffic delays.
- *Teamwork and interaction:* Air traffic controllers work in various team situations with team members within the ATC centre as well as outside the centre, including pilots and airport actors. Only when prolonged holding situations occur, a separate stack controller is assigned.
- *Procedures:* Procedures describe the formal or standard operating procedures for traffic handling. The morning inbound peak coincides the shift between night and day operation (determined by clocktime) and, dependent on the time of the year, the beginning of the daylight period. Different procedures exist for these variations, for example limitations in runway use during night-time operation or outside daylight conditions.
- *Technical systems:* Air traffic controllers use several systems to perform their tasks and generate information needed. This includes communication systems, planning systems, surveillance systems and decision support tools providing information and alerts to assist the ATCo.

Cognitive complexity issues in inbound peak operation

The characteristics of the operational situation, as described above for a typical inbound peak situation, result in cognitive complexity issues for the ATCo. Sources to identify complexity issues at LVNL are human factors analyses, interviews with controllers and bottlenecks found in acquiring expertise in ATC training at LVNL. The following categories of cognitive processes are distinguished in the ACoPOS model, for which complexity issues can be identified related to the inbound peak situation:

- *Situation assessment:* Situation assessment results in situation awareness, involving: 1) perceiving information, 2) interpreting the actual situation and 3) anticipating how the situation evolves. The complexity of situation assessment is created by, amongst others, frequent changes in the information that is perceived, continuous interpretation required of the solution space available and anticipating on emerging deviations between the actual and planned situation.
- *Attention management & workload management:* ATCos regulate their amount of attention and manage their workload depending on the specific situation. This includes monitoring the situation, directing attention to specific situations and keeping overview over the situation. The need for systematic scanning of the operational situation without being distracted by events and the frequent variations in workload require the ability to focus on specific situations, but also to be able to accelerate and extend the focus to multiple situations.
- *Problem solving & decision making:* ATCos solve problem- or conflict situations, formulate a plan for traffic handling and decide on what course of action to take. In a typical inbound peak, the deconfliction of bunches of traffic creates cognitive complexity. Standard solutions (e.g. speed and altitudes used) are applied, but switching to non-routine traffic handling is often required, increasing cognitive complexity.

- *Actions:* The outcome of the cognitive processes results in actions executed by the ATCo to interact with the operational environment. A busy inbound peak situation is typically characterised by a high RT load.

Air traffic controller strategies in disturbed ATC operation

To ensure safety whilst efficiency and environment are not sacrificed, controllers employ a combination of strategies, adjusted to the characteristics of the situation as well as operational constraints. Strategies reduce the likelihood of overall task performance being compromised (Histon & Hansman, 2008; Loft, Sanderson, Neal & Mooij, 2007; Malakis, Kontogiannis & Kirwan, 2010; Mogford et al., 1995; Nunes & Mogford, 2003). A strategy is defined as a working method or specific class of air traffic control activities that achieves one or more objectives (e.g. safety, orderliness, expeditiousness) within a certain investment of time and effort (Loft et al., 2007). Based on literature, a list of controller strategies has been generated, categorized into the cognitive processes in ACoPOS (see Schuver-van Blanken & van Merriënboer, 2012). However, the question remains which set of strategies are used by controllers in response to disturbed operational situations and whether other strategies exist in addition to those found in literature. Therefore, we started an exploratory study in 2012 focusing on the research question: *Which strategies do radar controllers use in response to disturbed ATC operation?*

To answer this question, retrospective interviews with individual expert controllers have been used, taking a disturbed operational situation they handled themselves as a basis (see Schuver-van Blanken & van Merriënboer, 2012). In addition to this approach, the method of focus groups was used to extract controller strategies, using the ACoPOS model as a basis. The results of a focus group on typical inbound peak operation in the morning as described in the previous paragraph are described below.

Focus group design

The ACoPOS model forms the basis for the structure of the focus group. 11 ACC controllers at LVNL participated, who are responsible for the ACC training of new air traffic controllers, with an average operational ATCo experience of 17 years. The focus group duration was 1,5 hours.

After an introduction on the purpose, a common mindset was created on typical inbound peak operation in the morning at ACC South sector. Four short movies with typical examples of operational traffic handling in an inbound peak in the ACC South sector were used for this. In addition, ACoPOS was used as a basis for a shared understanding of the (disturbing) factors present in typical inbound peak operation as well as to systematically address the cognitive processes involved. Next, expert insights were generated to get indications for strategies in typical inbound peak operation. This was done by probing questions structured around the ACoPOS cognitive processes, available on a large A3 paper as well as in PowerPoint. First, each participant wrote down their individual insights on the probing questions on the A3 paper. Then, each participant brought in their notes in a group discussion, where the insights of the group were collected in PowerPoint. The insights of the group covered both the answers on the probing questions, as well as the explanations on how they act and why.

Results of the focus group

Group results were categorised into answers that were identical or covered the same aspect or goal as well as the related ACoPOS cognitive processes. The categorised results were characterized by the underlying strategies that may apply. To do so, the explanations during the group discussion provided the necessary context for identification and characterization of the underlying strategy. Next, the individual results were analysed in the same way. The strategies that have been emphasized during the group discussion and were present in the individual results of at least 5 experts are presented in Figure 2, structured around the ACoPOS cognitive processes.

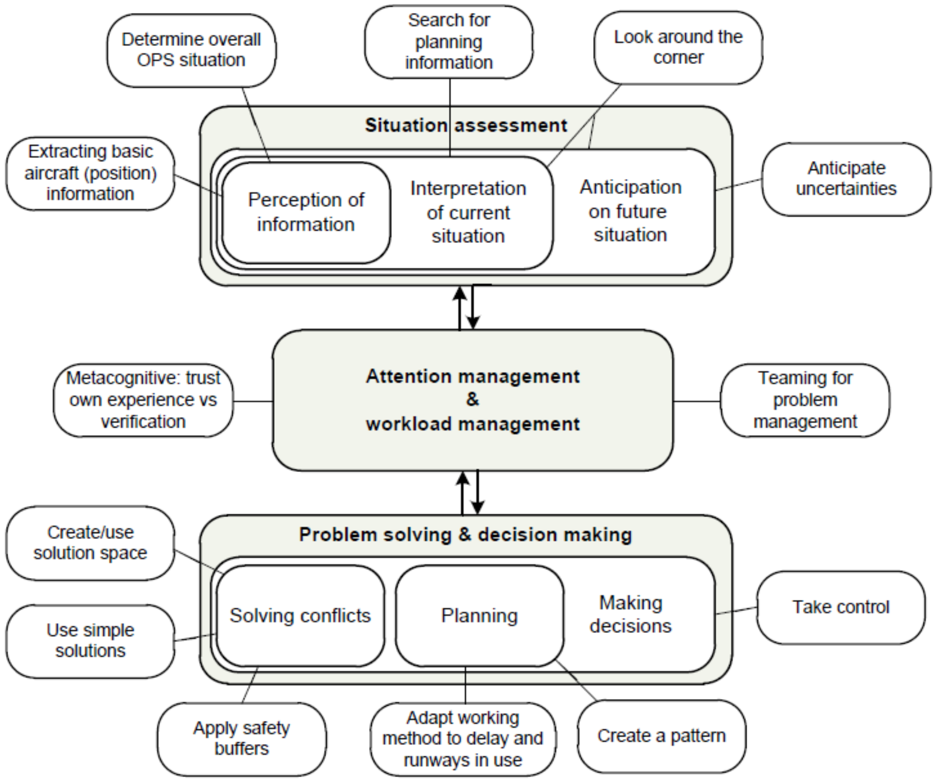


Figure 2, Air traffic controller strategies in inbound peak operation at the ACC South sector

In comparison to the list of strategies from literature (see Schuver-van Blanken & van Merriënboer, 2012), the results from the focus group indicate the presence of the following new strategies:

- *Determine the overall OPS situation*: The overall operational (OPS) situation is determined in a perspective being broader than the traffic situation. This includes weather, wind and visibility circumstances, the runways in use and airspace (un)availability.
- *Search for planning information*: Planning information is actively searched for, regarding (updates of) the expected approach time, the amount of delay or the expected inbound aircraft.
- *Look around the corner*: Controllers look around the corner to determine the traffic situation in the adjacent sector (e.g. traffic density or traffic handling) to be able to pro-actively act on this.
- *Metacognitive*: Results indicate that attention and workload is guided by a metacognitive strategy related to trusting one's own experience in judgment in a specific situation, versus verification of a potential problem situation using tools.
- *Teaming for problem management*: Teamwork with the adjacent sector or the planner controller is important to realize early or partial problem solving or to prevent problems.
- *Create/use solution space*: Controllers create solution space or use available space to solve problem situations. This is also done to prevent problems, to keep other solution possibilities available (e.g. to keep a vector possibility) or to maintain efficiency (e.g. for continuous descend).
- *Create a pattern*: In traffic handling, controllers create a pattern in their traffic handling, e.g. by creating a lateral pattern in vectoring or a structural buildup of the vertical pattern in holding operation. This also helps them to create overview and manage expectancies.

Indications for the strategies 'look around the corner', 'search for planning information', 'create/use solution space and 'create a pattern' were also found in the results of the retrospective interviews that focused on more complex disturbances (see Schuver-van Blanken & van Merriënboer, 2012).

Conclusions and recommendations

Clarifying what constitutes cognitive complexity in operational disturbances and which strategies are underlying ATCo expertise in these situations, enables us both to reduce cognitive complexity in ATC procedures and systems as well as to improve ATC training for acquiring ATCo expertise. By using the ACoPOS model as a framework, the factors influencing and causing cognitive complexity can be made visible. In addition, insights with respect to controller strategies to mitigate cognitive complexity can be systematically revealed. The results of the focus group revealed the strategies used by expert air traffic controllers in response to day to day disturbances in inbound peak operation in the morning in a dense area control airspace. New strategies were found in addition to literature and four of these strategies are in line with the results of the new strategies found in the retrospective interviews focusing on more complex disturbances in peak operation. This might indicate that the new strategies are crucial for mitigating daily disturbances in ATC peak operation. Analysis of additional cases, both by means of retrospective interviews as well as focus groups, is ongoing to determine which strategies are crucial in disturbed ATC operation, both in response to day to day disturbances as well as more complex disturbances.

References

Endsley, M.R. (1995). Toward a theory of situational awareness in dynamic systems. *Human Factors,* 37, 32-46.

Fothergill, S., & Neal, A. (2008). The effect of workload on conflict decision making strategies in air traffic control. In *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 39-43). New York.

Hilburn, B. (2004). *Cognitive complexity in air traffic control – A literature review.* Eurocontrol: EEC.

Histon, J.M. & Hansman, R.J. (2008). *Mitigating complexity in air traffic control: The role of structure-based abstractions.* ICAT-2008-05. Cambridge, MA: MIT.

Loft, S. Sanderson, P., Neal, A & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control. *Human Factors, 39* (3), 379-399.

Malakis, S., Kontogiannis, T. & Kirwan, B. (2010). Managing emergencies and abnormal situations in air traffic control (part I): taskwork strategies. *Applied Ergonomics, 41*(4), 620-7

Mogford, R.H., Guttman, J.A., Morrow, S.L., & Kopardekar, P. (1995). *The complexity construct in air traffic control: a review and synthesis of literature*. DOT/FAA/CT-TN95/22. Atlantic City International Airport, NJ: DOT/FAA Technical Centre.

Nunes, A., Mogford, R. H. (2003). Identifying controller strategies that support the ‘picture’. *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society.* Santa Monica, CA

Oprins E. (2008). *Design of a competence-based assessment system for air traffic control training*. Doctoral dissertation, University of Maastricht.

Oprins, E, Burggraaff, E. & van Weerdenburg, H. (2006). Design of a competence based assessment system for ATC training. *International Journal of Aviation Psychology, 16*(3), 297-320.

Redding. R.E., Ryder, J.M., Seamster,, T.L., Purcell, J.A. & Cannon, J.R. (1991). *Cognitive task analysis of en route air traffic control: Model extension and validation.*ERIC document No. ED 340 848. McLean, VA: Human Technology.

Schuver-van Blanken, M.J., Huisman, H. & Roerdink, M.I. (2010). The ATC cognitive process & operational situation model – a model for analysing cognitive complexity in ATC. *Proceedings of the 29th EAAP Conference,* Budapest, Hungary.

Schuver-van Blanken, M.J. & van Merriënboer, J.J.G. (2012). Air traffic controller strategies in operational disturbances - an exploratory study in radar control. *Proceedings of the 30th EAAP Conference*, Sardinia, Italy.

SESAR (2007). *Human Factors Impact.* SESAR Consortium: SESAR Definition Phase - Task deliverable WP1.7.1./D2.

EFFECTS OF AIRPORT TOWER CONTROLLER DECISION-SUPPORT TOOL ON CONTROLLERS' HEAD-UP TIME

Miwa Hayashi
NASA Ames Research Center
Moffett Field, California, USA

Jose M. Cruz Lopez
Universidad Autónoma de Nuevo León
San Nicolás de Los Garza, Nuevo León, Mexico

Visually monitoring aircraft traffic outside the window is an important part of tower controllers' tasks. Introducing new controller decision-support tools in the tower may increase or decrease their head-up time. The present study investigates the effect of NASA's Spot and Runway Departure Advisor (SARDA) tool on controllers' head-up time. A small video camera was mounted on controllers' heads to measure their head-movement activities during simulated tower operations with or without the SARDA advisories. Simple, pixel-color-based classification algorithms were able to achieve reasonably high accuracy of head-up detection in the training video frames (91% on average). The results showed that the controllers were looking outside 33% of the time, on average. The Local Controller looked outside about 8.5% more when the SARDA advisories were provided. No other major impact of the SARDA advisories on the head-up time was found in the study.

Air Traffic Control Tower (hereafter, *tower*) controllers are responsible for the separation of aircraft and the efficiency of traffic flow on the airport surface and in its designated airspace. Visually monitoring the traffic outside the tower window is an important part of fulfilling this responsibility. NASA is developing a tower controller decision-support tool called Spot and Runway Departure Advisor (SARDA) (Jung, et al., 2010), which provides scheduling advisories to the tower controllers. The present paper examines the effects of this tool on controllers' head-up time.

SARDA computes the runway-use sequence optimized in terms of taxi delay, Traffic Management Initiative constraint conformance, and runway throughput, and presents advisory information to controllers on their Electronic Flight Strip (EFS) system. More specifically, SARDA provides the Local Controller (LC), who controls the active runway traffic, with runway-use sequence advisories, and the Ground Controller (GC), who controls all of the other traffic in the Airport Movement Area, with departure metering advisories from the ramp area. One hypothesis of SARDA's impact is that the tool alleviates controllers' planning work and allows them to complete head-down tasks more quickly. As a result, the controllers can look outside more. A counter argument is that the tool demands more head-down time to read the advisories, and consequently reduces the head-up time. Every time the controller visually monitors the traffic outside, it adds an additional layer of protection against runway incursion and other traffic conflicts. Thus, any effects on controllers' head-up time have direct implications for the safety of the airport traffic management operation. To maintain safety, the proposed tool must not decrease controllers' head-up time.

Some researchers have investigated tower controllers' head-up and/or head-down time. Bos, Schuver-van Blanken, and Huisman (2012) reported that the controllers in their study were head-down for 74% of the time when they were using paper flight progress strips and 79% when they were using the EFS system in their tower operation simulation. Bos, et al. (2012) speculated that the increase of the head-down time may have been caused by additional visual attention required for moving electronic strips, and the head-down time might eventually decrease after the controllers became accustomed to handling the EFS. Pinska and Bourgois (2005) video-recorded the controllers at the Arlanda Airport tower in Sweden, and concluded that the controllers looked outside the window for about 30-40% of the time. Hilburn (2004) used an observer and a video camera to record tower controllers' activities at two major European aerodromes, and found that the controllers spent 43-49% of the time looking outside the window. Pavet (2001) also filmed the tower controllers in the South Tower at the Paris Charles de Gaulle airport, and reported that the controllers looked outside the window for about 20% of time. Together, these past studies suggest that, at busy airport towers, controllers generally spend about 20-50% of their time monitoring traffic outside the window.

The present study examines the effects of SARDA advisories on controllers' head-up time. Measurements were collected while the SARDA advisories were either provided to the controllers (*Advisory* runs) or not provided (*Baseline* runs). The EFS system was used in both types of runs. A small, lightweight wearable camera mounted on the side of the controller's head was used to record the controller's line of sight and later identify the controller's head-up time. The measurements were conducted during a one-week dress-rehearsal session of the main SARDA simulator evaluation session. The dress-rehearsal session was chosen for this data collection, rather than the actual main session, because of the unknown magnitude of the performance impact of mounting a camera on the controllers' heads. The dress-rehearsal session did not yield any performance results other than the head-up time measurement results reported in this paper. Readers interested in cross-referencing the current paper's results with the other performance results (e.g., workload ratings) are referred to the findings of the main evaluation session conducted in May 2012 with a similar simulation setup but with a different set of controller participants (Gupta, et al., 2013; Hayashi, et al., 2013). The results of the main session demonstrated that the use of SARDA significantly reduced the taxi delay, fuel consumption, and controller workload.

Method

Simulation

Facility. The simulation was conducted in the FutureFlight Central tower simulator facility at NASA Ames Research Center (Figure 1). The tower cab simulator provided a 360-degree computer-generated out-the-window view projected onto twelve 10-foot by 7-foot screens. The GC-position workstation consisted of a 24-inch touch screen monitor showing the EFS and a surface surveillance map monitor on the left side of the EFS monitor. The LC-position workstation included the same set of monitors as the GC workstation plus a third monitor on the right side of the EFS showing the airborne radar image.

Figure 1. FutureFlight Central Tower Simulator.

Traffic. The east-side traffic of the Dallas/Fort Worth International airport (DFW) in the south-flow configuration (runway 17R used for departures and 17C for arrivals) was simulated. Two levels of traffic, medium and heavy, were simulated, and two scenarios per each traffic level were generated for use. The medium-traffic scenarios consisted of 30 departures and 20 arrivals, and lasted for about 35 minutes. The heavy-traffic scenarios contained 40 departures and 27 arrivals, and lasted for about 45 minutes.

Experimental design. The measurements were collected in eight runs. Two retired tower controllers, who each had over 25 years of experience in DFW tower operations, participated in the simulation. The two participants were asked to perform the GC and LC tasks, alternating between these positions after each run. The four independent variables were Advisory (Advisory vs. Baseline runs), Position (GC vs. LC), Traffic Level (Medium vs. Heavy), and Participant (Controller 1 vs. 2). The runs were ordered to counterbalance potential learning or fatigue effects within and between each participant.

The dependent variables were the controllers' head-up behaviors, calculated from the video data using the methods described in the next section. Once their head-up times were computed, the head-up durations and frequencies were calculated and the effects of the SARDA advisories on these quantities, as well as the effects of the Position, Traffic Level, and Participant, were examined.

Video Data

Camera and video clips. The study used the Looxie 2 video camera (Looxie, Inc., Sunnyvale, CA) (Figure 2). This small, lightweight, wearable camera (3.25 inches long, 22 grams) was clipped onto the controller's one-ear headset,

Figure 2. Looxie 2 Video Camera.

positioned just above the ear not covered by the earphone, facing forward. Each controller wore one camera. The camera recorded a 480-pixel-by-320-pixel video clip at a rate of 15 frames per second. The recording format was MPEG-4. In this study, a 20-minute video-clip segment between the 5th minute and the 25th minute in each run, during which the traffic was sufficiently built up and the controllers were usually busy managing the traffic, was used for the analysis.

Classification. The Image Processing Toolbox of the MATLAB software package (release 2012b; MathWorks, Inc., Natick, MA) was used to extract frames from the recorded video and analyze the RGB triplet values of each pixel. After some trial and error, the following two simple classifiers shown in Table 1, or a combination of them, were found to perform the classification task sufficiently well:

Table 1. *Head-Up/Down Classifiers*

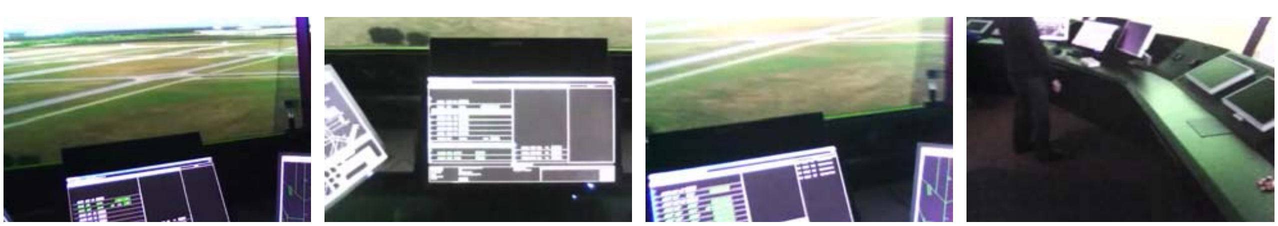
Classifier	Value	Classification
A	X_A = the sum of the RGB triplet values of all the pixels in the top half of the video frame	$X_A \geq \tau_A$: Head Up $X_A < \tau_A$: Head Down
B	X_B = the count of the black pixels in the top one-third of the video frame	$X_B \geq \tau_B$: Head Down $X_B < \tau_B$: Head Up

Classifier A utilizes the bright colors of the out-the-window image, and classifier B takes advantage of the dark color of the image inside the tower. A combination classifier AB first applies the classifier A to each frame, then the frames classified as Head Down are sent for further classification by classifier B. Which classifier works best differs for each video clip because of the camera angle. Classifier A works best if the camera was pointed slightly downward, and B works best if the camera was facing slightly upward. For some clips, combination classifier AB works the best.

The optimal threshold for each classifier (τ_A, τ_B) also varies among the clips. To obtain the best threshold for each classifier for a certain clip, and then the best classifier for the clip, the following training process was performed for each clip:

1. Extract video frames at a 3-second interval. A 20-minute clip generates 400 such frames. These serve as a training set.
2. Manually label each of the training frames as *Up*, *Down*, *Ambiguous* (i.e., it is hard to judge Up or Down), or *Other* (i.e., the controller looked at a different place). See Figure 3 for examples.
3. Run each of the classifiers, A and B, with a varying threshold on the training frames labeled Up or Down. Find the best threshold (τ_A, τ_B) that resulted in the maximum match rate between the classification result and the labels placed in Step 2. Run the combination classifier, AB, with the best thresholds for A and B.
4. Select the classifier among the three that resulted in the maximum match rate.

Figure 4 shows an example of the selection of the best threshold (Step 3). In this example of classifier A, a threshold for the sum of the RGB triplet values, $\tau_A = 1.8 \times 10^7$, yields the highest match rate—93%. A graph like Figure 4 also provides some information about the sensitivity of the match rates to the selection of a particular threshold. If the peak of the curve is relatively flat, the selection affects the match rate little. If it is sharp, the match rate is more sensitive to the selection. In the example in Figure 4, the curve exhibits a relatively flat peak. Any



(a) Up (b) Down (c) Ambiguous (d) Other

Figure 3. Examples of Manually Labeling the Training Frames.

threshold between 1.6×10^7 and 2.0×10^7 would result in a match rate over 90%, which would be reasonably close to the highest rate of 93%. After performing the above four steps, all video clips resulted in a match rate of 91% on average (range: 83-98%), which was sufficiently accurate for the goals of this study.

Once the classifier and the threshold for each clip were selected, the head-up versus head-down classification could be applied automatically to the remaining frames in the clip. In this study, frames at an interval of 0.25 seconds were processed, rather than all 15 frames per each second, assuming that this sampling rate would be fast enough for identifying human head movements. The frames labeled Other in Step 2 and any frames within ±1.5 seconds of these frames were labeled Other in the classification to prevent them from mixing into either the Up or Down category.

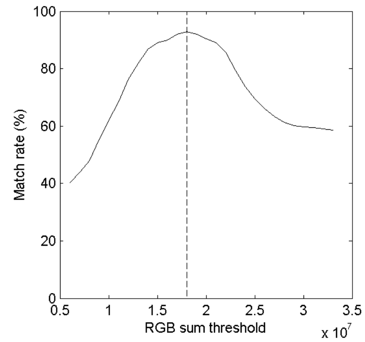


Figure 4. An Example of the Selection of the Best Classifier Threshold.

Results

Head-Up Behavior Statistics

Table 2 lists the statistics of the controllers' head-up behaviors. *Std Dv* is the standard deviation, and *N* is the number of observations used to compute the mean and the standard deviation.

- The % Head-Up Time column shows the statistics of the percentage of the time in the entire length of the 20-minute segment that the controllers were looking outside in the applicable video-clip group.
- The Head-Up Duration (sec) column shows the statistics of the duration of out-the-window viewing observed in the applicable video clips. (The distributions of these values are heavily skewed to the right.)
- The Head-Up Frequency Per Minute column shows the statistics of how frequently the controllers looked outside within each of the 20 one-minute segments in the individual video clip.

Table 2. *Head-Up Behavior Statistics*

Video-Clip Group	% Head-Up Time			Head-Up Duration (sec)			Head-Up Frequency Per Minute		
	Mean	Std Dv	N	Mean	Std Dv	N	Mean	Std Dv	N
Advisory run	34.3%	9.5%	8	2.6	3.3	1254	7.8	2.4	160
Baseline run	31.5%	11.5%	8	2.7	3.3	1130	7.1	2.7	160
Medium traffic	35.1%	11.9%	8	2.8	3.5	1187	7.4	2.6	160
Heavy traffic	30.6%	8.5%	8	2.5	3.1	1197	7.5	2.6	160
Ground	38.6%	6.3%	8	3.3	3.9	1116	7.0	2.3	160
Local	27.1%	10.5%	8	2.1	2.5	1268	7.9	2.7	160
Controller 1	28.7%	9.6%	8	2.4	3.0	1166	7.3	2.7	160
Controller 2	37.1%	9.7%	8	2.9	3.5	1218	7.6	2.4	160
All video clips	32.9%	10.3%	16	2.7	3.3	2384	7.5	2.6	320

Analysis of Variance

The head-up frequencies per one-minute segment, whose means are listed in Table 2, were subjected to a mixed-model analysis of variance (ANOVA). The main effects included were Advisory, Position, Traffic Level, and Participant. In addition, the three two-way interaction effects involving Advisory effect (i.e., Advisory × Position, Advisory × Traffic-Level, and Advisory × Participant) were included in the model. Participant effect was treated as a random effect, and all the other effects were treated as fixed effects (Lindman, 1974). The results showed that

Participant effect was statistically significant ($F_{1,304} = 19.4$, $p < 0.001$). No other effect was found statistically significant.

Analogously, the same ANOVA model was applied to the total head-up duration per one-minute segment. The analysis found a statistically significant Participant main effect and an Advisory × Position interaction effect ($F_{1,304} = 14.3$, $p < 0.001$; $F_{1,1} = 3218$, $p = 0.01$; respectively) and a marginally significant Advisory × Participant interaction effect ($F_{1,304} = 3.7$, $p = 0.057$). Figures 5 and 6 plot the means to visualize the directions and magnitudes of these effects. (Error bars are omitted because the statistical significances were already inferred by the ANOVA.) Figure 5 shows that the total head-up duration per minute increased in the LC position when the SARDA advisories were provided (13.7 sec in Baseline, 18.8 sec in Advisory), but not in the GC position (24.1 sec in Baseline, 22.3 sec in Advisory). Figure 6 indicates that, for Controller 1, the total head-up duration per minute increased when the SARDA advisories were presented (19.6 sec in Baseline, 22.9 sec in Advisory), but not for Controller 2 (18.1 sec in Baseline, 18.2 sec in Advisory).

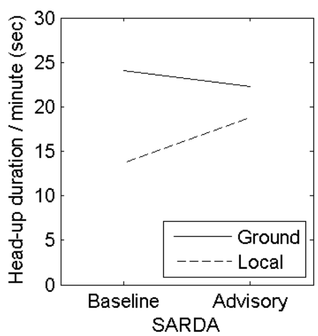


Figure 5. Means of Total Head-Up Duration Per Minute by Advisory × Position.

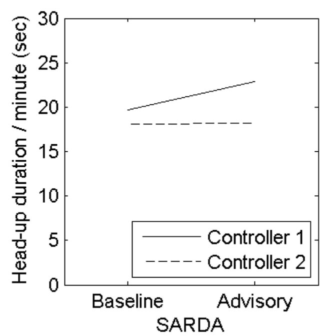


Figure 6. Means of Total Head-Up Duration Per Minute by Advisory × Participant.

Discussion

Table 2 indicates that the controllers were looking outside the window for 32.9% of the time on average. The ±2σ-range was 12% to 53% (σ = 10.3%), which roughly overlaps with the previous studies' findings of 20-50% of head-up time. The % Head-Up Time was about 3% higher in the Advisory runs than the Baseline runs, but the difference may be too small to draw a reliable conclusion. Indeed, the ANOVA results did not find any statistical significance in Advisory main effect on either the head-up frequency or duration.

The ANOVA did find that the head-up durations were significantly increased in the Local position when the SARDA advisories were provided, but not in the Ground position (Figure 5). Table 2 suggests that, the % Head-Up Time is lower in the Local position than in the Ground position in general. It also suggests the LC makes more frequent and shorter glances at the out-the-window view than the GC does. These results together suggest that both GC and LC look outside just as frequently with or without the SARDA advisories, but that SARDA advisories allow the LC to monitor outside for a longer time (about 5 sec longer, or 8.5% more, per minute), probably because the advisories allow the LC's head-down tasks to be completed more quickly.

Aside from the above Advisory × Position effect, overall, the SARDA advisories had little impact on the controllers' head-up time, at least in the simulation setup used in this study. Surprisingly, the effect of Traffic Level was not found significant, either. The Heavy traffic level used in the simulation may not have been high enough, relative to the Medium traffic level used, to cause any detectable differences in the controllers' performance. The effects of the controllers' personal biases, which appeared as the significant Participant main effect and Advisory × Participant interaction effect, were found most prominent in the head-up behavior data. That means that individual controller may have very different head-up behaviors, which may also be affected differently by SARDA.

Lastly, the study employed a low-cost, commercial, off-the-shelf video camera and simple pixel-color thresholds to identify the controllers' head-up time. This method is not the most accurate way to measure controllers' scanning activities, but it allowed for fast setup and attained reasonably high classification accuracy (91% match rates in the training set) well suited to the goals of this study. A weakness of this method is that some frames were ambiguous as to whether the controller was looking up or down. In this study, to decrease the ambiguous frames as much as possible, the controllers were asked to be standing instead of sitting during the runs. This increased the required vertical movement of the head between looking up and down, and facilitated the classification of head-up vs. head-down to a certain extent. Still, about 10% of the training frames fell into the Ambiguous category. This is a limitation of not only the current study's method, but also any head-movement-based classification method. Researchers who are interested in higher accuracy are advised to use more sophisticated tracking systems, such as eye trackers. The other shortcoming of the current approach was the inability to detect sideways head motions. When the controllers looked sideways to look outside through a different window, the classification might have resulted erroneously in Head Down because the lower part of the window frames in the distance rises in a 3D perspective, making the intensity level of the frame similar to those of the Head Down frames (Figure 4d illustrates a similar situation). Also, the current method could not distinguish if the controller was looking at the EFS or the map display. Advanced object recognition techniques could be employed to overcome these shortcomings.

Conclusion

The study found that the controllers looked outside for about 33% of the time during the simulated tower operations, a result consistent with previous studies' findings. The results showed that the SARDA advisories helped the Local Controller to look outside for about 5 seconds or 8.5% more per each minute. No other major impact of the SARDA advisories on the head-up time was found in the study. Large Participant effects observed in the head-up behavior measurements suggest that, to obtain more generalizable results, it is essential for future research to include a wider range of controller participants.

References

Bos, T. J. J., Schuver-van Blanken, M., & Huisman, H. (2012). *Towards a paperless air traffic control tower* (Report No. NLR-TP-2011-192), Amsterdam, the Netherlands: National Aerospace Laboratory NLR.

Gupta, G., Malik, W., Tobias, L, Jung, Y., Hoang, T., & Hayashi. M. (2013, submitted). Performance evaluation of individual aircraft-based advisory concept for surface management. *Tenth USA/Europe Seminar on Air Traffic Management Research and Development*, Chicago, IL.

Hayashi, M., Hoang, T., Jung, Y. C., Gupta, G., Malik, W., & Dulchinos, V. L. (2013, submitted). Usability evaluation of the Spot and Runway Departure Advisor (SARDA) concept in a Dallas/Fort Worth airport tower simulation. *Tenth USA/Europe Seminar on Air Traffic Management Research and Development*, Chicago, IL.

Hilburn, B. (2004). *Head down time in aerodrome operations: a scope study*, Technical Report, Den Haag, The Netherlands: Center for Human Performance Research.

Jung, Y. C., Hoang, T., Montoya, J., Gupta, G., Malik, W., & Tobias, L. (2010). A concept and implementation of optimized operations of airport surface traffic. *Tenth AIAA Aviation Technology, Integration, and Operations (ATIO) Conference.* Fort Worth, TX.

Lindman, H. R. (1974). *Analysis of variance in complex experimental designs.* San Francisco, CA: W. H. Freeman and Company.

Pavet, D. (2001). Use of paper strips by tower air traffic controllers and promises offered by new design techniques on user interface. *Fourth USA/Europe Air Traffic Management Research and Development Seminar*, Santa Fe, NM.

Pinska, E., & Bourgois, M. (2005). Behavioural analysis of tower controller activity, *EUROCONTROL Experimental Centre, Activity Report 2005*, Brétigny-sur-Orge, France: EUROCONTROL.

STRUCTURING, ANALYZING AND MONITORING PROBLEMS AND DECISION MAKING PROCESSES AT CIVIL AIR NAVIGATION SETS OF A PUBLIC ORGANIZATION

Cabral, Lisia Maria Espinola da Silva Pacheco
Brazilian Airport Administration Organization - INFRAERO / Av. 20 de Janeiro, s / nº., Antonio Carlos Jobim International Airport - Galeão, Ilha do Governador, CEP 21941-570, Rio de Janeiro, RJ, Brazil, Phone: 55-21-33984310, 55-21-99569228, E-mail: lisiacabral@ig.com.br
Estellita Lins, Marcos Pereira
Problem Structuring and Indicators Group for Modeling and Assessment - PSIGMA, Rio de Janeiro Federal University - UFRJ / Av. Horácio Macedo, nº. 2.030 / F-110, Fundão, Ilha do Governador, CEP 21.941-914, Rio de Janeiro, RJ, Brazil, Phone: 55-21- 2562-7047, E-mail: estellita@pep.ufrj.br

This article presents the partial application of a post-graduation doctoral study developed at civil Air Navigation sets of a public organization, on operational work stations: Tower; APP – Approach; and AIS - Aeronautical Information System Room. Its main goal is to demonstrate how Concept Map was used as a Multi-methodology and an instrument of SODA - Strategic Options and Development Analysis, in SOR - Soft-Operational Research, for PSM - Problem Structuring Methods, to contribute for structuring, analyzing and monitoring problems and decision making processes, by optimizing interdisciplinary “iteractions” and interactions to: promote System and Rational Thinking; reduce self-deception and increase metacognition; encourage Metagovernance and Situational Leadership; enable possible improvements. Its main purpose is to consolidate a predictive culture on Operational Safety Management, according to aeronautical requirements, focusing on Human Factors, for the continuously emergence of hidden threats, considering Complex System´s characteristics of complexity, variability, dynamics, unpredictability and minimum acceptable risk control.

Introduction

The search of human being for modern technologies in response to social and organizational demands represents a challenge. Nevertheless, there is a reductionist tend towards quantifying information with standardized models, putting a risk into its reflection and effectiveness. Therefore, is required to complement this approach with a more refined analysis to explicit knowledge for better deal with unexpected situations and future foresee. There for, it was used Concept Map in this study as a Multi-methodology (GHARAJEDAGHI, 2011) and an instrument of SODA - Strategic Options and Development Analysis, in SOR - Soft-Operational Research, for PSM - Problem Structuring Methods (ESTELLITA, 2010; ROSENHEAD and MINGERS, 2001), basing the main subject of this article, that is structuring, analyzing and monitoring problems and decision making processes in Complex Systems.

I. Conceptual Base

Concept Map (ESTELLITA, 2010; ARÊAS, 2009) was used in this study as a Multi-methodology (MINGERS, 2006) and an instrument of SODA - Strategic Options and Development Analysis, in SOR - Soft-Operational Research, for PSM - Problem Structuring Method (ESTELLITA, 2010; ROSENHEAD and MINGERS, 2001), to promote the reflection about different and conflicted perceptions of the same problems, by interdisciplinary interactions (GHARAJEDAGHI, 2011) and Metacognition (FLAVELL, 1976), for better understanding the reality. In the same way, Cognitive Map provides mental representations about the reality, but it was not used in this study.

Multi-methodologies (MINGERS, 2006) consist of a number of new methodologies used to address problems involving material, social and individual qualitative aspects, once just one methodology is not complete enough for this. This study used Concept Map (ESTELLITA, 2010; ARÊAS, 2009) as a Multi-methodology to encourage the practice of thinking in group in order to create a space out of work routine to provide reflection, enlarging the chances of better understanding its strengths and limits.

This study found on System Thinking (GHARAJEDAGHI, 2011) and Rational Thinking (SENGE, 2008) a form to define more precisely problems, solutions and possible qualitative changes, based on: power dimensions; knowledge; well-being; beauty; values; and desire. For this, it is necessary successive “iteractions” involving human behavior and continuous interactions among different actors, providing new alternatives and desirable objectives for the future (GHARAJEDAGHI, 2011). Concept Map (ESTELLITA, 2010; ARÊAS, 2009) was used in this study as a Multi-methodology (MINGERS, 2006) to consolidate the "iterative loopings" and a better understanding about the reality of the work stations studied.

According to Metacognition Theory (FLAVELL, 1976), the state of mind depends on: knowledge dimensions - propositional, experimental, representative and epistemological, that interfere on perception; alternatives to deal with the levels of consciousness; elements analysis for the construction of references to pass from the subjective to the inter-subjective attribute; characterization and measurement of the relevant subjective attributes for decision making process (MINGERS, 2006). Also, the "Unknown Unknowns" (YAMIN, 1286-1368) shows how people choose for self-deception to be accepted by a group, by fear of going against the status-quo, assuming bad attitudes on disrespect of others instead of promoting positive changes to maintain their integrity. This study demonstrates how Metacognition (FLAVELL, 1976) may help and self-deception (YAMIN, 1286-1368) may harm the commitment to face problems.

Governance is the transcendence of an isolated part of an integrated whole, by interactions among various organizations in order to regulate what they have in common, emphasizing the centralization or the localization of the organizational power. Metagovernance (JESSOP, 2002) assumes the interdependence of relations and question hierarchic coordination values of all levels (top-down / bottom-up) for a more reflexive organization, by the commitment with effectiveness of: economic controls; collective objectives; and associated values. As participants of this study, there are operators and their heads, including supervisors, coordinators and managers, adopting a free communication channel with dialogue ("requisite variety") that enables to analyze the same situation from different perspectives. If not opened to other level´s points of view, the failures of the high levels of the organization hierarchy will remain static.

Situational Leadership (HERSEY & BLANCHARD, 1986) describes two kinds of maturity for the operator´s heads deal with the staff at work: psychological maturity, involving relationship; and task maturity, involving knowledge, ability and experience. This study associates Situational Leadership to Metacognition (FLAVELL, 1976) and Metagovernance (JESSOP, 2002), to better deal with reality.

Complexity is one of the main Complex System´s characteristics (ESTELLITA, 2010), as retreated in the Air Navigation sets of the public organization studied, and it is not consistent because of the main paradoxes that involve these kind of systems, as follows: individual and collective object; isolated and global interdependent parts; centralized and distributed information; individual and multiple observers; conservation and transcendence of processes. This will be commented in this article.

II. Method

This study used a method based on SODA - Strategic Options and Development Analysis, in SOR - Soft-Operational Research, for PSM - Problem Structuring Method (ESTELLITA, 2010; ROSENHEAD and MINGERS, 2001), which the main elements are, as follows:

2.1. Scenario – Operational work stations (Tower, APP – Approach, AIS - Aeronautical Information System Room) of six Civil Air Navigation sets of a public organization, and one will be commented here.
2.2. Participants – Operators of Air Traffic Control (Tower and APP – Approach); AIS - Aeronautical Information System, Meteorology and Air Navigation Communication (AIS Room); and operators´ heads (supervisors, coordinators, managers).
2.3. Study Focus – Structure, analyze and monitor problems and decision making processes in Air Navigation Complex Systems.

III.Methodology

Concept Map (ESTELLITA, 2010; ARÊAS, 2009) was applied during this study as a Multi-methodology (MINGERS, 2006), preceded and succeeded of others, what embodied four phases, as follows:

3.1. 1st. Phase / 2011 - Problems Representation: In the year of 2011, a group exercise was conducted and, after the preliminary definition of a representative for each group, it was split into seven stages:

3.1.1. Brainstorm – Each group was asked to select a problem of the addressing operational set and, by the use of brainstorm, register, randomly, the ideas associated, on one side of a poster paper, with a pilot pen.
3.1.2. Symbolization – Each group was asked to analyze the ideas linked to the problem selected and register, now with symbols (map, flowchart, tree-diagram, bubble, design, image etc.), its characteristics, effects, possible solutions and necessary interventions, from a systemic point of view, on the other side of the same poster paper already used, with a pilot pen.

3.1.3. Simulation – Each group was asked to simulate a real situation involving the problem selected and debated (brainstorm / symbolization), to demonstrate it, in practice, what, after approval, was filmed.
3.1.4. Oral presentation – Each group was asked to make an oral presentation of the problem selected and debated (brainstorm / symbolization), and its implications, based on the stages above, with a conclusion.

3.2. 2nd. Phase / 2011 - Problems Consolidation: It was consolidated all the problems and implications, based on the stages of the 1st. Phase / 2011 - Problems Representation, what embodies three stages:

3.2.1. First Concept Map (ESTELLITA, 2010; ARÊAS, 2009) – The First Concept Map (ESTELLITA, 2010; ARÊAS, 2009) was elaborated for each operational set, from the first "iterative looping" (GHARAJEDAGHI, 2011) conducted during the 1st. Phase / 2011 - Problems Representation, containing the consolidation of all the problems once addressed.
3.2.2. First report – A written report was elaborated for each operational set, containing the problems once addressed during the 1st. Phase / 2011 - Problems Representation, and the suggestions of interventions, what was sent to each local operators´ heads and the higher levels of managers.
3.2.3. Debriefing –An oral debriefing about the results (First Concept Map - ESTELLITA, 2010; ARÊAS, 2009 / First Report) was conducted to the local operators´ heads and the higher levels of managers.

3.3. 3rd. Phase / 2012 - Problems Representation Update: Another group exercise was conducted, based on the First Concept Map (ESTELLITA, 2010; ARÊAS, 2009), elaborated during the 1st. Phase / 2011 - Problems Representation, and, after the preliminary definition of a representative for each group, it was split into two stages:

3.3.1. Debate – A debate followed by a written update of the First Concept Map (ESTELLITA, 2010; ARÊAS, 2009) was accomplished.
3.3.2. Oral presentation - Each group was asked to make an oral presentation with a conclusion, based on the debate above.

3.4. 4th. Phase / 2012 - Problems Consolidation Update: It was consolidated all the problems´ update and their implications, based on the stages of the 3rd. Phase / 2012 - Problems Representation Update, what embodies two phases:

3.4.1. Second Concept Map (ESTELLITA, 2010; ARÊAS, 2009) - The Second Concept Map (ESTELLITA, 2010; ARÊAS, 2009) was elaborated for each operational set, from the second "iterative looping" (GHARAJEDAGHI, 2011) conducted during the 3st. Phase / 2012 – Problems Representation Update, containing the consolidation of all the problems addressed before.
3.4.2. Second report – A written report was elaborated for each operational set, containing the problems addressed before during the 3st. Phase / 2012 – Problems Representation Update, and the suggestions of interventions, what was sent to each local operators´ heads and the higher levels of managers.

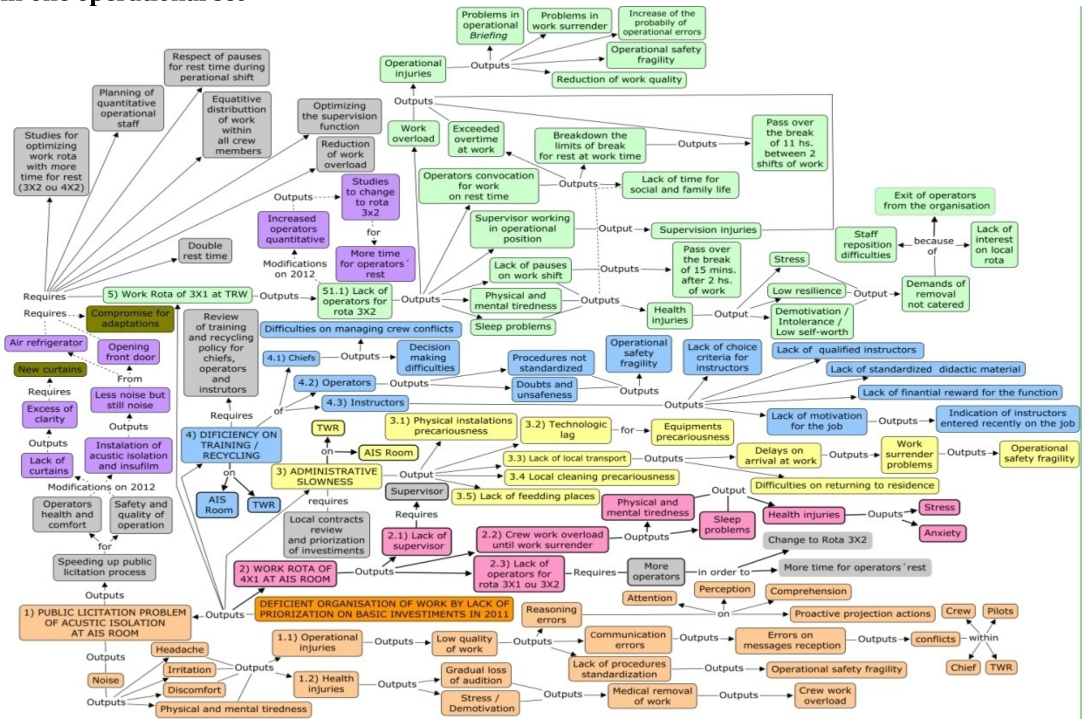
So, in 2011, the present study accomplished the first "iterative looping" (GHARAJEDAGHI, 2011) with the 1rd. Phase / 2011 - Problems Representation, followed by the First Concept Map (ESTELLITA, 2010; ARÊAS, 2009) with the 2th. Phase / 2011 - Problems Consolidation; and, in 2012, the study proceeded with the second "iterative looping" (GHARAJEDAGHI, 2011) with the 3rd. Phase / 2012 - Problems Representation Update, followed by the Second Concept Map (ESTELLITA, 2010; ARÊAS, 2009) with the 4th. Phase / 2012 - Problems Consolidation Update.

This article is restricted to comment the 1st. Phase / 2011 - Problems Representation and the 2nd. Phase / 2012 - Problems Consolidation of only one operational set studied, leaving the 3rd. Phase / 2012 - Problems Representation Update and the 4th. Phase / 2012 - Problems Consolidation Update for the future.

IV. Analysis

"Figure 1" shows the First Concept Map (ESTELLITA, 2010; ARÊAS, 2009) for one operational set, related to the 2nd. Phase / 2012 - Problems Consolidation, and represents a Multi-methodology (MINGERS, 2006), because it is part of a sequence of others: it is a result of the first "iterative looping" (GHARAJEDAGHI, 2011), related to the 1st. Phase / 2011 - Problems Representation; and also served to the continuity of the subsequences phases – the second "iterative looping" (GHARAJEDAGHI, 2011), related to the 3rd. Phase / 2012 - Problems Representation Update, and the Second Concept Map (ESTELLITA, 2010; ARÊAS, 2009) related to the 4th. Phase / 2012 - Problems Consolidation Update.

"Figure 1": First Concept Map (ESTELLITA, 2010; ARÊAS, 2009) with the problems consolidation in one operational set



According to "Figure" 1, this First Concept Map (ESTELLITA, 2010; ARÊAS, 2009) content, related to the 2nd. Phase / 2012 – Problems Consolidation, begins with a central problem – "Deficient Organization of Work by Lack of Priorization in Basic Investiments in 2011", with five main developments:

(1) Public Licitation of Acustic Isolation – The Law 8666 / 1993 - Licitation for Public Organizations, prescribes that the winner company of a licitation is the one that presents the lowest price for the service, what may damage its quality. The current licitation of acustic isolation for AIS Room lasted too much time in consequence of an unsuccessful licitation involving an unconcluded service by the company first licitated, what obliged to open a new licitation. On the other hand, this law represents a protection for not accepting any company to do the services needed without licitation, and this exemplifies one of the many paradoxes of Complex System (ESTELLITA, 2010) we can find in the public organization studied, equivalent to the need of conservation this process, but also to transcend it for the necessary improvements.
(2) Work Rota of 4x1 at AIS Room – This problem involves: lack of supervisor; crew work overload until work surrender; and lack of operators for rota 3 x1 or 3 x2. It indicates that some operational staff was transferred to other operational sets without reposition. There was also less operational staff because of medical removals, vacations, absences, without having enough agility in personal reposition or mechanisms to minimize the negative impacts of this reality. These causes of staff reduction may be minimized for short periods of time, but it is a self-deception (YAMIN, 1286-1368) to deny the real demand for more operators. The plan for calculating operators quantification must include this type of occurrences, but public organizations, as the one studied, tend towards presenting bureaucratic routines, retarding the solutions for this problem, what is an obstacle to change work rota, so this aspect needs to be improved.
(3) Administrative slowness (AIS Room and Tower) – Slow and bureaucratic processes involving; physical installations precariousness; technologic lag; lack of local transport; local cleaning precariousness; lack of feeding places etc., demonstrate that public organizations, as the one studied, are not generated enough to prepare themselves to make the necessary interactions to attend common objectives. This characterize another paradox of Complex System involving isolated and global interdependent parts (ESTELLITA, 2010): these processes need to be addressed as a global interdependent part of a whole, to have the agility enough to be solved, and not to remain only as an isolated part separate of the whole. Therefore, they remain slow and bureaucratic faced to real demands, so this aspect also needs to be improved.
(4) Deficiency on Training / Recycling (AIS Room and Tower) – This problem embodies the lack of a career plan including a periodicity of defined courses for each operational function (AIS Room and Tower),

operators´ heads and instructors, restraining them to up-grad towards the advance of aeronautic technology. The training plan is restricted to formative courses, although there was an investment on recycling courses (English Language and Operational Safety), what increased the staff knowledge. This represents another paradox of Complex System involving centralized and distributed information (ESTELLITA, 2010), where the lack of training indicates a trend towards centralized information and the investment on training indicates a trend towards distributed information, once even more information people get, more knowledge and safety they have, what is the base for an aimed predictive culture (CANADA, 2006, 2010, 2012).
(5) ***Work Rota of 3x1 at Tower*** – This problem involves the lack of operators for rota 3 x 2, indicating a trend towards: work overload; operators convocation for work on rest time; supervisor working on operational position; lack of pauses on work shift; physical and mental tiredness; and sleep problems. See the same considerations of problem (2), agravated by the stressful activity of air traffic control at Tower.

The problems described above have their respective subdevelopments showed in "Figure 1". In general, public organizations have difficulty to prioritize solutions in response to the existent demands and, in consequence, do not succeed on making continuous qualitative changes required by Complex System´s characteristics, as complexity, variability, dinamics, unpredictability and acceptable risk control, in complement to normative prescription (CANADA, 2006, 2010, 2012). Therefore, in one year (2011 - 2012), there were neither the necessary improvements for the problems addressed, nor the necessary predictions of new problems to appear, for all the operational sets, not attending the expectation of more agility on this.

Nevertheless, both of the group exercises conducted during the study, either the first one, related to the 1st. Phase / 2011 - Problems Representation, or the second one, related to the 3rd. Phase / 2012 - Problems Representation Update, were considered useful by a majority of participants, what demonstrates the significance of the "iteractions" provided by these two successive "iterative loopings" (GHARAJEDAGHI, 2011), enabling the elaboration of the Concept Maps (ESTELLITA, 2010; ARÊAS, 2009) with the problems addressed. Besides, both exercises provided the opportunity to make interactions (GHARAJEDAGHI, 2011) and treat conflicts among different professional formations of different hierarchical levels, what enlarged the chances to: minimize self-deception (YAMIN, 1286-1368) enhance Metacognition (FLAVELL, 1976); put into practice Systemic Thinking (GHARAJEDAGHI, 2011) and Rational Thinking (SENGE, 2008); apply Metagovernance (JESSOP, 2002) and Situational Leadership (HERSEY & BLANCHARD, 1986); and increase the understanding about Complex System and its characteristics involving complexity (ESTELLITA, 2010), and the commitment of each person and groups with the appropriate management of problems and their risks. Comparing both exercises, we have: during the first one, related to the 1st. Phase / 2011 – Problems Representation, a majority of participants from all hierarchic levels demonstrated a certain motivation to face problems and hope to solve them; but during the second one, related to the 3rd. Phase / 2012 – Problems Representation Update, the participants evoked a certain frustration for the slowness of problems conduction and resolution, although unintentionally.

Also, there is a prevalence of bureaucratic characteristics (CANADA, 2006, 2010, 2012) in the organization studied, what, through Complex System´s (ESTELLITA, 2010) point of view, indicates the relevance to understand them for making them treatable. Related to the dynamics of the paradoxes of Complex System (ESTELLITA, 2010), we have that: anyone´s individual objects show up more than everyone´s collective common objects, although these last ones are desirable, for enabling to situate the organization members in different contexts; isolated parts are more evident than the interdependent whole, characterized by multifunctional ambiguity in different areas of knowledge, once different sectors of the same operational set and different operational sets lean towards functioning in a separated and independent way, not having the habit to consider the necessities among each other or to supply them interactively; power centralization is more present than the necessary systemic distribution of information among different observers, based on an integrated vision of the whole, in which the communication among different operational sets, sectors and hierarchical levels, not always has the same priority; individual unity is emphasized in relation to multiplicity of different observers, what requires a continuous observation of reality, from different points of view, in diverse, similar, conflicted or complementary activities, instead of a shy knowledge exchange about internal and external occurrences, as observed in the study; processes preservation are prevalent in relation to their transcendence, what could bring continuous and permanent evolutions, subjected to emergent properties of Complex System´s self-organization, once there is a trend towards the routine preservation, based on the current regulations, and not on their constant update.

In reference to Metagovernance (JESSOP, 2002), it is reinforced a reciprocal interdependence of crescent and frequent interactions - intra and inter-organizational, intra and inter-sectional, and national and international. But the partial results of this study shows that there are evidences of a confuse, distant and slow communication in the organizational hierarchy, that needs to be continuously treatable and improved for its safety, based on the concepts of Complexity (ESTELLITA, 2010). Related to Situational Leadership (HERSEY & BLANCHARD, 1986), there is a trend of the operators´ heads towards not making a difference between the task maturity and the psychological maturity in their operational performance, what may difficult the appropriate management of the staff in their work stations, deviating from the main focus on people to the focus on processes management, what contributes to the extension of self-deception (YAMIN, 1286-1368) instead of Metacognition (FLAVELL, 1976). This suggests to intensify the necessary continuation of "iteractions" and interactions (GHARAJEDAGHI, 2011), using Concept Maps (ESTELLITA, 2010; ARÊAS, 2009); and to stimulate Metagovernance (JESSOP, 2002) and Situational Leadership (HERSEY & BLANCHARD, 1986), to improve communication and better cope with reality.

V. Study Application

The present study is still being developed in its 4th. Phase / 2012 - Problems Consolidation Update, and have resulted into this article and another poster presentation at SHE – Sustainability, Health and Education Conference / 2012, by UFRJ – Rio de Janeiro Federal University: "A Qualitative Methodology for Problems Structuring, Analyzing and Monitoring Applied to a Complex Public Organization".

VI. Conclusion

This study demonstrates that is possible to structure, analyze and monitor problems and decision making processes, using Concept Map (ESTELLITA, 2010; ARÊAS, 2009) as a Multi-methodology (MINGERS, 2006) and an instrument of SODA - Strategic Options and Development Analysis, in SOR - Soft-Operational Research, for PSM - Problem Structuring Methods (ESTELLITA, 2010; ROSENHEAD and MINGERS, 2001), based on: Systemic Thinking (GHARAJEDAGHI, 2011) and Rational Thinking (SENGE, 2008); Metacognition (FLAVELL, 1976) and self-deception (YAMIN, 1286-1368); Metagovernance (JESSOP, 2002); Situational Leadership (HERSEY & BLANCHARD, 1986); Complexity and Complex System (ESTELLITA, 2010). Therefore, it aims at contributing for transforming the organizational culture in more predictive and generative to improve the effectiveness of Operational Safety Management (CANADA, 2006, 2010, 2012).

VII. Future Prospective

It is intended to be written more literature about this study theme and, in the future, advance from the application of Concept Map to Cognitive Map, as Thinking Map (ESTELLITA, 2010; ARÊAS, 2009).

VIII. Acknowledgements

Thanks to the public organization for the trust to participate of the study; my husband and family for the respect of time dedicated on it; my friends for the patience to share difficulties; my post-doctoral advisor for the opportunity to develop it; and my post-doctoral colleagues for the exchange of knowledge.

References

-CANADA (2006, 2010, 2012). ICAO (International Civil Aviation Organization). Doc 9859 / AN 474: Safety Management Manual. In: http://www.icao.net/publications.
-ESTELLITA LINS, Marcos Pereira; ANTOUN NETTO, Sérgio Orlando; SAMANEZ BISSO, Cláudio R. (2010). Coursepack - Complex social problems: Structuring with Concept Maps. Post-graduation Course in Production Engineering of COPPE / UFRJ.
-FLAVELL, John H. (1976). Metacognitive aspects of problem solving. *In L. B. Resnick (Ed.)*, The nature of intelligence (pp. 231- 235), New Jersey, NJ: Lawrence Erlbaum.
-GHARAJEDAGHI, Jamshid (2011). Chapter 7: Design thinking, Book: System thinking, Managing the chaos and complexity: A platform for designing business architecture. Ed. Elsevier, USA.
-HERSEY, Paul. & BLANCHARD, Ken H. (1986). Theory and techniques of situational leadership. São Paulo: EPU.
-JESSOP, Bob. (2002). Governance and metagovernance: On reflexivity, requisite variety and requisite irony. Lancaster University in http://comp.lancs.ac.uk/sociology/soc108rj.htm.
-MINGERS, John (2006). Chapter 3 - Living systems: Autopoiese, Book: Accomplishing thinking systems: Knowledge and action of management. Ed. Springer, University of Canterbury, UK.
-ROSENHEAD, Jonathan; MINGERS, John (2001). A new paradigm of análisis. *In: Rosenhead, J.; Mingers, J. (eds) Rational analysis for a problematic world revisited.* Chichester, John Wiley & Sons.
-SENGE, Peter (2008). The fifth discipline. Ed. Best Seller.
-YAMIN, Ibn (1286-1368).Unknown Unknowns. *Adapted by GSE (1979).*

OUTCOME-BASED RISK PATHWAYS: UTILIZING SAFETY REPORTS TO UNDERSTAND RISKS IN AIR TRAFFIC CONTROL

Katherine A. Berry, Michael W. Sawyer, & Edward M. Austrian
Fort Hill Group, LLC
Washington, DC

Many accident investigation taxonomies have been developed over the years to assist in identifying and classifying causal factors and errors involved in near misses events and accidents. While these taxonomies are often used to better understand individual events, they also offer the potential for quantifying the relationships between causal factors and errors to better understand emerging systemic issues. In an effort to extend beyond traditional frequency-based accident analysis, this work details the relationships among causal factors by examining any differences among outcome types. An analysis of 417 Aviation Safety Reporting System (ASRS) reports yielded five key risk pathways present in air traffic control safety events involving actual or near losses of standard separation minima. These risk pathways allow mitigation strategies to be targeted directly at the key causal factors in order to produce the greatest positive impact on the system as a whole.

The ability to identify and understand human performance safety trends is necessary in complex industries, such as air traffic control (ATC). Furthermore, the development of a baseline of human performance for current day operations has the potential to impact the designs of future systems. Providing designers with knowledge of the current human performance trends in a system permits designers to incorporate mitigations aimed at those trends into the earliest stages of concept development (GAO, 2011). The human factors safety community of practice has long served as a key player in the enhancement of safety in the ATC domain. Human factors causal factors assessments are typically conducted using criteria, such as calendar year, domain type, geographic region, meteorological conditions, and many other conditions (Sawyer, Berry, & Austrian, 2012). Safety events, however, are seldom an outcome of one single causal factor, but are more commonly the culmination of multiple, related factors (Senders & Moray, 1991).

While many studies identify leading causal factors in frequency-based assessments, little has been completed to examine the relationships among the various causal factors within the air traffic domain. It is beneficial to the safety community to expand beyond traditional frequency-based assessments to incorporate causal factor relationship assessment. Furthermore, the examination of risk pathways that identify and quantify the statistically significant relationships among causal factors should be expanded to include event outcomes. The development and implementation of mitigations strategies based only on the most frequent error types has historically proven difficult due to the variability associated with human performance (Berry, Stringfellow, & Shappell, 2010). The associations determined by the risk pathways approach will assist in driving mitigations upstream. Since the higher-tier causal factors (e.g., agency influences) are associated with less variability, mitigation strategies targeted at these latent conditions may have the potential to produce "the greatest gains in safety benefits" (Li & Harris, 2006). Establishing risk pathways will aid in driving mitigation strategies targeted towards latent conditions while still incorporating active errors.

Purpose

This assessment presents the results of an analysis of safety event reports submitted by air traffic controllers describing safety events observed in live National Airspace System (NAS) operations. By analyzing reports of actual operations, the work presents a picture of the human factors safety issues associated with varying outcomes in NAS operations from an ATC perspective. In order to achieve this purpose, a customized air traffic safety taxonomy was developed based on an analysis and synthesis of existing taxonomies including Human Factors Analysis Classification System (HFACS) (Wiegmann & Shappell, 2003), JANUS (Isaac et al., 2003), and HERA (Isaac et al., 2003). The Air Traffic Analysis and Classification System (AirTracs) taxonomy was then applied to examine the underlying trends present in 417 ATC safety events resulting in a near or actual loss of standard separation minima (LoSS) with any significant differences between near and actual LoSS findings being examined. The prominent risk pathways among the AirTracs causal categories were identified. These identified human performance trends should serve as a foundation of current day human performance operational knowledge for human factors practitioners and NextGen system designers in the early stages of concept development.

Methodology

AirTracs provides a framework for systematically and thoroughly examining the impact of human performance on air traffic accidents and incidents. The framework of the AirTracs causal category model is based on the Department of Defense (DoD) HFACS model (DoD, 2005), while the detailed causal factors incorporate factors from HERA and JANUS (Isaac et al., 2003). The AirTracs framework promotes the identification of causal trends by allowing factors from the immediate operator context to agency-wide influences to be traced to individual events. The causal category model is displayed in Figure 1. For more information on the AirTracs causal factor categories see Sawyer, Berry, & Austrian, 2012.

The data utilized for this analysis was gathered from NASA's Aviation Safety Reporting System (ASRS), which is comprised of voluntarily submitted aviation safety reports filed by pilots, controllers, or other NAS actors (NASA, 2013). As with any voluntary reporting system, ASRS combines the advantages of direct input on safety concerns from front-line personnel with the disadvantages of potentially biased points-of-view. For this study, ASRS safety incidents resulting in a near or actual LoSS event were queried for reports filed by an air traffic controller and occurring in the calendar year 2011. The resulting 417 ASRS reports were classified with AirTracs utilizing the consensus method, which required a consensus or agreement on the causal factors contributing to the report by a panel. The panel members included human factors experts, retired air traffic controllers, and flight deck experts. Each report was evaluated across all levels of the AirTracs framework, and the presence or absence of each AirTracs causal category was recorded. It is important to note that the AirTracs categories are not mutually exclusive. For example, an individual report can include both an execution act and a decision act.

A Person's chi-square test or Fisher's exact test were utilized to identify any statistical differences among AirTracs causal categories when comparing near LoSS and LoSS events. If any statistical differences were identified between the two outcomes, the relative risk value for the significant causal category was calculated. For those causal categories that did not result in any significant differences between near LoSS and LoSS events, the risk pathways or associations among causal categories were examined. Starting at the highest AirTracs tier Agency Influences, the relationship between each causal category at the higher tier and the various causal categories at lower tiers was examined using a Pearson's chi-square test to measure the statistical strength of the association. In the instances where the assumptions of the Pearson's chi-square test were not met, a Fisher's exact test was conducted (Sheskin, 2011). If the AirTracs category resulted in a significant association being identified through the Pearson's chi-square test or Fisher's exact test ($p<0.05$), the odd's ratio was calculated for that particular association. The odd's ratio is a measure of the degree of the association strength that compares the odds of the presence of causal category (Sheskin, 2011).

Results

The findings from the AirTracs analysis of 417 near LoSS or LoSS ASRS reports can be viewed in Figure 1. The percentages in Figure 1 do not sum to 100% since reports typically are associated with more than one causal factor. When examining the differences between the findings for LoSS reports and near LoSS reports, only the Cognitive/Physiological causal category produces significant differences between LoSS and near LoSS reports ($X^2 = 5.8212$, $p<0.05$, Relative Risk=1.4600). Since only one significant difference was found between LoSS and near LoSS reports, the risk pathways were examined in the aggregate for all causal categories except for Cognitive/Physiological pairings.

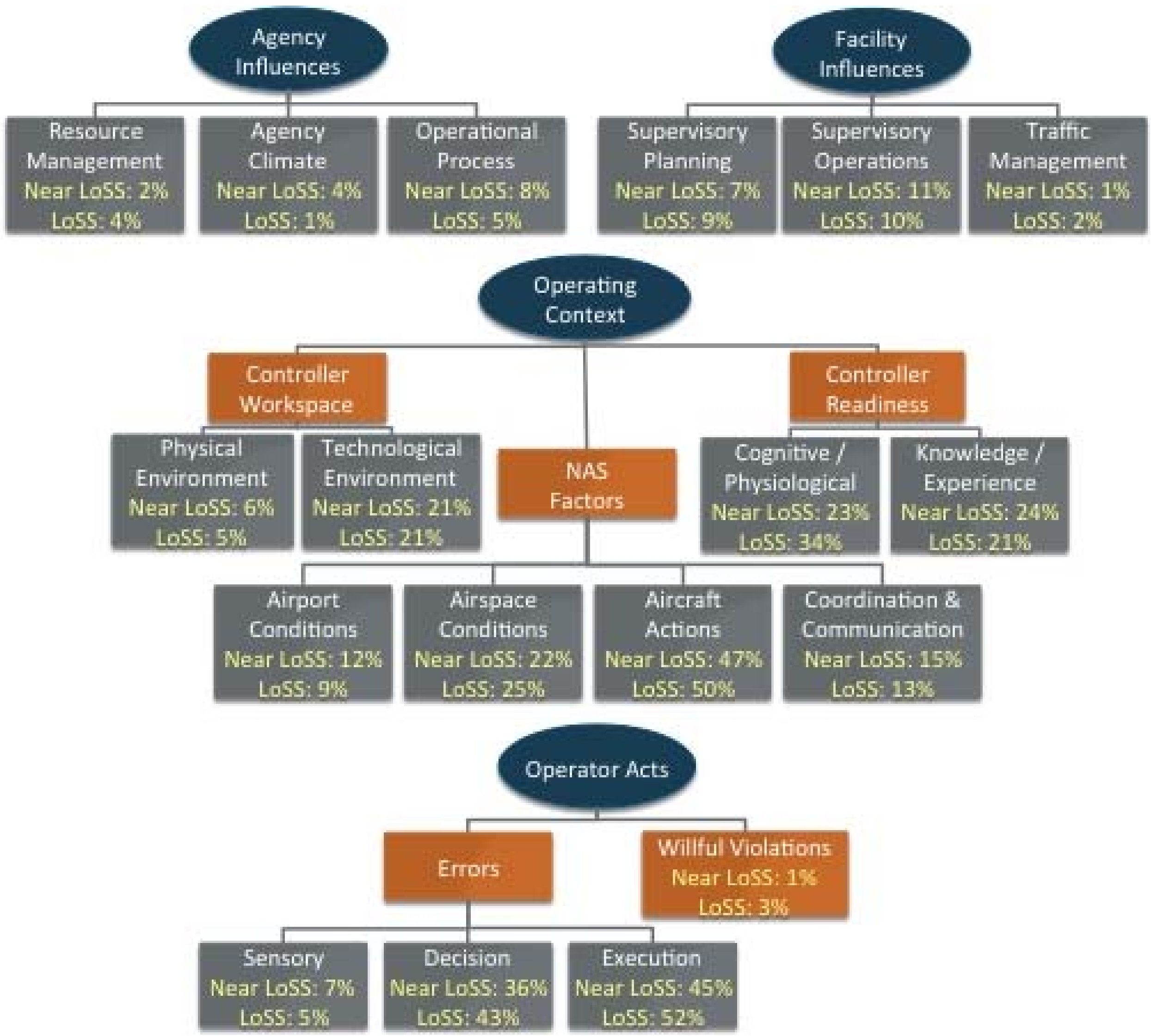


Figure 1. AirTracs Findings (Percentage of ASRS report with causal category presence indicated for each outcome)

The AirTracs risk pathways where statistically significant associations between causal categories were found are shown in Table 1. Only those causal category pairings that were found significant from the Pearson's Chi-Square analysis and odds ratio analysis ($p<0.05$) were reported. Five risk pathways were identified containing ten causal category associations. When the LoSS and near LoSS reports were compared, the Cognitive/Physiological causal category resulted in significant differences between the two outcomes. When examining Cognitive/Physiological pairings, one significant association emerged for each outcome. For LoSS outcome reports, the Cognitive/Physiological causal category was significantly associated with Execution Errors ($X^2 = 5.5640$, $p<0.05$, OR = 1.9542), and for near LoSS outcome reports, the Cognitive/Physiological causal category was also significantly associated with Execution Errors ($X^2 = 8.1969$, $p<0.05$, OR = 2.6600).

Table 1. LOS and Near LOS Risk Pathways

AirTracs Causal Categories	Pearson's Chi Square	Odds Ratio
Resource Management Pathway		
Resource Management X Supervisory Planning	5.8103	4.7000
Supervisory Planning X Technological Environment	9.4869	3.0358
Technological Environment X Sensory Error	13.7832	4.4911
Aircraft Actions Pathway		
Aircraft Actions X Decision Error	12.1495	2.0291
Aircraft Actions X Execution Error	20.0748	2.4438
Airport or Airspace Condition Pathways		
Supervisory Planning X Airport Condition	4.3151	2.5249
Supervisory Operations X Airspace Condition	12.3069	3.0154
Traffic Management X Airspace Condition	8.9855	13.5319
Knowledge/Experience Pathway		
Knowledge/Experience X Decision Error	4.2199	1.6212
Sensory Error Pathway		
Physical Environment X Sensory Error	22.5564	8.9176

Note. All Pearson's Chi Square values and Odds Ratios presented above are significant ($p<0.05$).

Discussion

When examining the LoSS reports versus the near LoSS reports, only the Cognitive/Physiological category resulted in significant differences between the two outcomes. The findings indicated that LoSS reports were 1.4600 times more likely to include a Cognitive/Physiological causal category than near LoSS reports. Sample causal factors within the Cognitive/Physiological causal category include high workload, attention, complacency/boredom, automation reliance, fatigue, expectation bias, and medical illness. In situations where separation could potentially be lost, this finding indicates that when a controller is under pressure from mental or physical conditions, a resulting outcome of an actual LoSS versus a near LoSS is more likely. These Cognitive/Physiological factors may either greatly contribute to the complexity of the event or may inhibit the controller from adequately detection and preventing the event.

The lack of significant differences among the remaining causal categories indicates that contributing factors may be similar regardless of outcome when comparing LoSS versus near LoSS events. This finding may indicate that controller's errors and mistakes are similar regardless of the outcome (near or actual LoSS). A controller may make an incorrect decision or poorly execute a plan leading to either a near or actual LoSS. However, if a manager is attempting to reduce the number of LoSS events by transforming the LoSS events into near LoSS events, managers should not focus on those causal categories that lack significant differences (e.g., execution error) and should focus on the causal category that did result in a significant difference – Cognitive/Physiological factors.

Risk Pathways

From the assessment of the aggregated LoSS and near LoSS reports, five risk pathways were identified that included ten causal category associations. The five pathways will be discussed in the following sections.

Risk Management Pathways. The risk management pathway is initiated with an association between the Risk Management causal category and the Supervisory Planning causal category, followed by an association between the Supervisory Planning causal category and the Technological Environment causal category, and resolved with an association between the Technological Environment causal category and the Sensory Error causal category. This connection illustrates how agency-level decisions regarding resources, such as budget, personnel, or equipment, can create a ripple effect throughout the causal chain and among various levels of the organization. Furthermore, the association between Technological Environment and Sensory Errors indicates that LoSS or near LoSS events with a Sensory Error may be attributed to the controller's technology or workstation not being salient enough to alert the controller to a potential conflict.

Aircraft Actions Pathway. The aircraft actions pathway incorporates the Aircraft Actions causal category associations with both the Decision Error and Execution Error causal categories. This pathway suggests that weaknesses exist among the decision/response selection and response execution phases of information processing, rather than the perception phase. Therefore, mitigations should be targeted towards improving controller's decision-making and plan execution in response to unexpected aircraft actions, such as pilot deviations and go-arounds. In 2013, the FAA (Teixeira, 2013) identified the current fiscal year's top five hazards in the NAS with one of the top five hazards being recovery. The aircraft actions pathway coincides with the recovery top five hazard. Both the hazards and the pathway identify the manner in which controllers respond to adverse events as being a current shortcoming in the NAS. Mitigation strategies for the recovery hazard should aim to address both decision errors and execution errors.

Airport or Airspace Conditions Pathways. The airport or airspace conditions pathways incorporate either the Airport Conditions causal category or the Airspace Conditions causal category. Both of these causal categories describe the design and environmental conditions of either the airspace or airport the aircraft is operating in, such as weather or sector/airport layout. The airport or airspace conditions pathways also include the Supervisory Planning, Supervisory Operations, and Traffic Management causal categories. This pathway suggest that the plans and actions of facility management, such as front line managers and traffic management unit, impact the way aircraft operate in and controllers respond to the environment and conditions of the airspace or airport. For example, if a traffic manager does not issue a traffic management initiative to reroute aircraft around a weather system, the aircraft or a stream of aircraft may encounter a weather system causing the controller to respond to the adverse condition under a more time-sensitive situation. Further analysis and research should be conducted to identify the details of the errors and actions of those actors at the facility management level.

Knowledge/Experience Pathway. The knowledge/experience pathway includes the Knowledge/Experience causal category and the Decision Error causal category. This pathway suggests that if a controller lacks experience with a situation (e.g., an unfamiliar procedure) or knowledge of a certain task (e.g., a controller in training), the controller's decisions, choices, and plans may not be adequate for the situation. Mitigation strategies should incorporate improving the knowledge base of the controller, which may be achieved with the least adverse impact to the safety of the NAS by simulator-based training. Controllers should complete simulator scenarios and conditions that may be unfamiliar or precarious to improve the decision-making process.

Sensory Error Pathway. In addition to the Sensory Error – Technical Environment association, the Sensory Error causal category was also associated with the Physical Environment causal category. These two association pairing illustrate the importance of the workspace conditions on the controller's perception. For example, if the control room is too noisy or the radar screen not bright enough, the conditions can impact the controller's ability to detect a potential conflict.

Conclusions

In order to examine the dynamic relationships of causal factors, an expansive human factors taxonomy, AirTracs, was utilized to identify prominent risk pathways based on a particular outcome. The

AirTracs taxonomy was utilized in assessing 417 ASRS air traffic control reports that resulted in a near or actual LoSS. The percentage of reports linked to each causal category was identified for both outcomes. The AirTracs outcome findings were compared and only the Cognitive/Physiological causal category resulted in a significant difference. For both outcomes, the Cognitive/ Physiological causal category was associated with the Execution Error causal category. When combining the near and actual LoSS reports, five key risk pathways were identified, and potential mitigation strategies were discussed. Targeting systemic mitigation strategies offers the potential to proactively reduce risks associated with the causal factors within the pathway.

References

Berry, K., Stringfellow, P., & Shappell, S. (2010). Examining error pathways: an analysis of contributing factors using HFACS in non-aviation industries. In *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*, 2010, San Francisco, CA.

Department of Defense (2005). *DOD HFACS: A Mishap Investigation And Data Analysis Tool*. Retrieved 2011 from http://www.public.navy.mil/navsafecen/Documents/aviation/aeromedical/DoD_hfacs.pdf.

Government Accountability Office. (2011). *Aviation Safety: Enhanced Oversight and Improved Availability of Risk-Based Data Could Further Improve Safety* (GAO-12-24). Retrieved from http://www.gao.gov/products/GAO-12-24.

Isaac, A., Shorrock, S.T., Kennedy, R., Kirwan, B., Anderson, H., & Bove, T. (2003). *The Human Error In ATM Technique (HERA-JANUS).* (EUROCONTROL Doc HRS/HSP-002-REP-03).

Li, W. C., & Harris, D. (2006). Pilot error and its relationship with higher organizational levels: HFACS analysis of 523 accidents. *Aviation, Space, and Environmental Medicine*, 77(10), 1056-1061.

NASA. (2013). *Aviation Safety Reporting System*. Retrieved from http://asrs.arc.nasa.gov.

Sawyer, M., Berry, K., & Austrian, E. (2012). The use of odds ratios and relative risk to quantify systemic risk pathways in air traffic control. In the *Proceedings of the International Society of Air Safety Investigators 2012 Annual Seminar*, 2012, Baltimore, MD.

Senders, J. W., & Moray, N. P. (1991). *Human Error: Cause, Prediction, and Reduction*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Sheskin, D. J. (2011). *Handbook of Parametric and Nonparametric Statistical Procedures* (5th ed.). Boca Raton: Chapman & Hall.

Teixeira, J. (2013). *Improving Aviation Safety: An Air Traffic Control Perspective* [PowerPoint Slides]. Retrieved from https://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/safety/

Wiegmann, D. A., & Shappell, S. A. (2003). *A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System*. Burlington, VT: Ashgate Publishing, Ltd.

MEASURING HUMAN FACTORS SUCCESS IN ACQUISITIONS

Jason Demagalski / Dr Branka Subotić
FAA / NATS
Washington DC, USA / London, United Kingdom

Dr Jerry Crutchfield / Dr Dwayne Norris
FAA / American Institutes for Research
Oklahoma City, USA / Washington DC, USA

Human factors specialists work as part of multidisciplinary teams supporting acquisition programs. Depending upon the acquisition model and culture of an organization, the exact points of involvement of human factors specialists may vary, as too will the metrics of success for each project. Acceptable/optimal level of operational performance is the goal of any Air Traffic Control (ATC) system. Human performance is a key component of operational performance. Measuring human performance can be an area of academic debate often yielding vague answers or need for more research. This is of little use or interest to project and senior managers who make decisions. This paper presents a high level overview of the difficulties of measuring performance, provides a case study of how this has been addressed in the United Kingdom ATC organization and concludes with thoughts on how the human factors community can support acquisitions and achieve what we would deem, success.

Human factors specialists work as part of multidisciplinary teams with a variety of stakeholders within acquisition programs and projects. Depending upon the organization and its acquisition structure, the exact stage/s and magnitude of involvement may vary. These inputs may range from the initial concept development stage through to operational implementation or merely an opportunity to "sign-off" at a set point in the acquisition process when anything but acceptance would be unwelcomed. While everyone involved in a project is seeking success, it is not sufficient to assume that all involved have the same criteria for success.

There are many intended and unintended consequences to an ambiguous definition of project success. If we ensure that the project success criteria include valid and verifiable measures of human performance then the chance of consulting with and incorporating the design recommendations from human factors specialists up front will be enhanced. The extent to which this can be achieved will reflect the dominant measures of success for the ATC organization. The ultimate goal of human factors specialists is for auditable and traceable human factors inputs to track from early requirement definition to final validation of human performance.

The central problem for human factors specialists lies in what organizations call success in acquisition programs. Is success when the hardware is installed? When the software functions correctly? When the operator can use the new tool/equipment? Ultimately, success must be when the overall system performance, as a result of the acquired system or tool, meets given criteria, one of which being that the human operator's task performance is within the acceptable range. If an organization is striving for an acceptable level of system and task performance following any change, then success of the project should be based on achieving it.

Within ATC, operational performance can be measured using a variety of metrics, such as traffic throughput, safety statistics and system availability. Different stakeholders will judge their success against the metric most closely aligned with their specialism. However, within large ATC organizations, there is considerable risk that the stove-piped nature of development of tools, technology, airspace changes, safety initiatives and such like, when implemented, lead to a resultant performance level less than that of the individual components. As such, if acquisitions and operational implementations are poorly managed strategically, then the Air Traffic Controller Specialist (ATCS) is left to individually integrate their knowledge and understanding of each newly acquired tool/procedure to allow them to perform effectively. Projects may be deemed to have been successful upon installation and activation of the new tool in the ATC facility regardless of how the controllers overall performance is affected.

In the last 20 years the pace of development of potential new ATC systems and the opportunities for their implementation have increased. As a consequence, this has increased the burden on organizational acquisition which in turn has increased the importance of having useful human factors requirements, the importance of testing and validating proposed systems and the importance of having consistent and reliable measures to indicate whether a system has fulfilled the requirements and is fit for purpose.

A key component of being able to measure human performance is to have defined performance standards. Performance standards describe the expected level of performance, in measurable terms, on critical work tasks. Developing reliable and valid performance standards requires the identification of the current work requirements. Work requirements are typically comprised of some combination of the tasks that are performed on the job; the knowledge, skills, abilities, and other attributes (KSAOs) required to perform those tasks; and the tools and equipment incumbents used to perform those tasks. Such requirements are identified through a job analysis and the resulting requirements can be used to support a variety of activities from personnel practices, acquisition requirements and establishing performance standards.

Performance standards define the level of performance required or expected at a given level of expertise. While the job analysis identifies the relevant work tasks, performance standards capture how the tasks are done or their expected results. Said another way, task lists allow organizations to differentiate between jobs; performance standards allow organizations to differentiate degrees of controller proficiency/performance on a specific job task or group of tasks. Performance standards should be realistic or challenging, specific, measurable, consistent with organization goals, and understandable. Performance standards can be used to develop assessments for a variety of purposes, or for setting the cut scores for those assessments. In addition, they provide a way to communicate the expectations of the organization. The level of performance at which the standards are built should be based upon the purpose for which the standards will be used.

There are a number of ways to set performance standards. A common method for setting performance standards, subjective standard setting, while the simplest, can also be the least effective. Using this method, standards are set without reference to criteria for effective performance. Instead, they are set arbitrarily, typically based on convention or rules of thumb (e.g., 70% is a passing score on a test). Norm-referenced methods are based on performance relative to others; for instance, training instructors might decide that the top 50% of a given training class has high enough performance levels, while the bottom 50% might need additional training. Criterion-referenced methods are a different and more thoughtful way of setting performance standards. These methods require that standards are based on specific criteria that are decided in advance (e.g., mastery of a work tasks at specified level of performance). Given the parameters within ATC, performance standards defined by ATCSs are ideally established using a criterion-referenced approach.

Within the criterion-referenced method, there are two general procedures that can be used to establish performance standards. Judgmental procedures require SMEs to make judgments about how work tasks are to be performed with respect to a target level of performance (e.g., the level of performance for a minimally acceptable candidate). Alternatively, empirical procedures establish performance standards at a level that differentiates task performance between two groups that vary in their level of qualifications (e.g., contrasting experts and novices). Both methods are acceptable for setting criterion-referenced performance standards. Krokos, Baker, Norris, and Smith (2007), developed performance standards for ATCS in tower, Terminal Radar Approach Control (TRACON), and Enroute environments. An example can be found in table 1.

Table 1. Enroute Control Center Performance Standards: Sample

Example Tasks	**Performance Standards.** The Successful Controller...…
1. Project mentally an aircraft's future position 2. Ensure separation using data from various sources of information including range/bearing function, vector lines, route function, or User Request Evaluation Tool (URET) function etc. 3. Determine whether aircraft/airspace separation standards may be violated 4. Review radar display to ensure aircraft compliance with clearance	**Behavioral Indicators of Certified Professional Controller (CPC)-level Performance:** • Appropriately uses tools to observe alerts, have vector lines out to watch projected traffic • Uses the Range Bearing Indicator (in URET) to measure distance • Initiates control actions prior to being alerted to a potential conflict **Cognitive Indicators of CPC-level Performance:** • Takes wider scan in order to understand how control actions affect the aircraft around them now and in the future • Develops multiple potential plans of action to resolve a potential conflict • Projects future position of aircraft **Results of CPC-level Performance in this Sub-Activity:** • No conflict alerts occur, and no "immediate" or "expedite" type control instructions (e.g. turn right immediately) are issued. • Surrounding controllers are not negatively affected by controller's actions. • All aircraft are separated safely.

Performance standards describe the expected level of performance on each activity at a given level of expertise. It is important to remember that these performance standards are not measurement tools. However, the standards could be used to develop such tools, thus allowing them to realize their full potential. For example, knowledge tests, job simulations, performance rating scales, or checklists could be developed to measure performance based on the standards.

Efforts to identify a set of generally accepted performance measures for air traffic control human factors evaluations have been plagued by challenges for decades (Hopkin, 1980). Challenges include access to data (due to technical feasibility, resources needed to collect the data, etc.), reliability of subjective criteria and the relationship of subjective criteria to objective criteria, as well as the largely cognitive, and therefore unobservable, nature of the air traffic control task. Additionally, these data are complex and dynamic in nature which can lead to analytical challenges, for example, comparing results collected from one type of sector to another type of sector or within the same sector but with different air traffic situations. Continuing advances in human factors methodologies and in information technology have over time presumably mitigated some of the challenges. With the increasing tendency of proposals to enhance the air traffic control system to include wider and more far reaching changes, it seems as important as ever to identify a set of standard performance measures so that the results of evaluations of proposed changes to tools and procedures can be consistent and comparable. It may be time once again for our community to examine the performance measures available and try to come to a consensus on what the performance measures should be.

Even with improvements in general methods and greater access to data, the challenges unique to the air traffic control domain still remain. New technologies and procedures that are evaluated in different ways and by different organizations in the industry only exacerbate these unique challenges. A standard set of scenario events and conditions would help to put results from different studies in context with relation to each other. Events would likely include off-nominal occurrences (e.g. convective weather, radio outage, runway shut-down) and conditions would likely be based upon predictions of traffic levels and complexity at certain milestones in the future. An application of our understanding of air traffic and air space complexity to a method to compare human performance experiencing different air traffic or using different air space may prove feasible and beneficial as well.

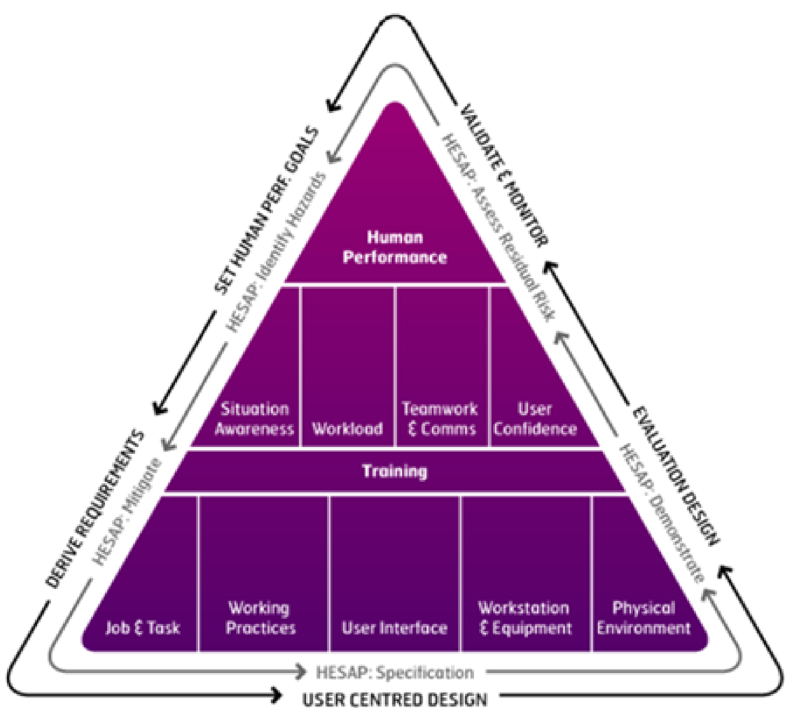
Human factors design through to performance in an international aviation organization (NATS, UK)

NATS is the United Kingdom's (UK) largest air traffic navigation service provider. NATS provides services to aircraft flying through UK controlled airspace and at several UK and international airports. In 2010/11, NATS operational staff handled 2.1m flights (87% of flights to/from UK, 13% overflights; NATS corporate presentation, 2012).

Interim Future Area Control Tools Support (iFACTS) for one of the NATS Area Control Centre (ACC) was born out of Research and Development (R&D) driven innovation that matured over years into a set of control tools. iFACTS represents a set of controller tools that use the current position of the aircraft, the flight plan and complex mathematics to predict the likely future position of all the aircraft in a sector. These positions are then compared for up to 18 minute into the future and any potential conflicts are highlighted to the controller. The most noticeable change from the previous system is the removal of the paper flight progress strips previously used by the controllers, and the subsequent requirement to enter all tactical clearances into the system electronically.

iFACTS completed transition into service in November 2011. On this particular project, the NATS Human Factors (HF) team was involved from design stages, through implementation, to monitoring of the system in live operation within its first six months of operation. All HF activities conducted on iFACTS were structured using the NATS HF assurance framework (Figure 1). This framework is split into emergent properties (depicting how the whole system performs) and design characteristics (describing characteristics of the individual parts of the system).

Figure 1. Human Factors Assurance Framework



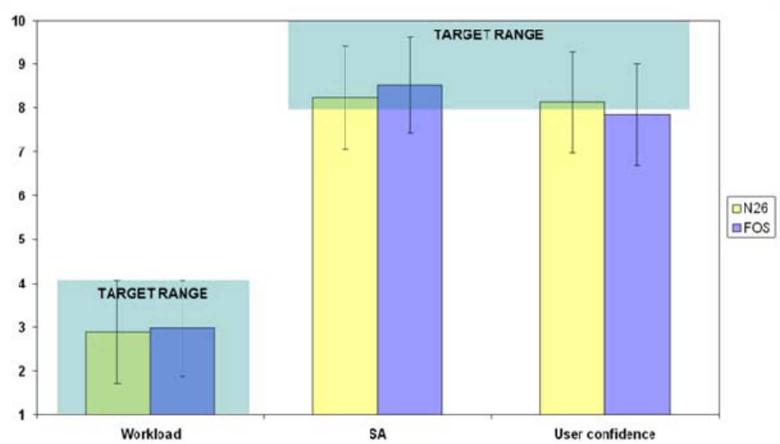
At the bottom of the triangle lie design characteristics describing individual parts of the system. The design characteristics can be directly designed. Therefore, we can specify requirements and objectives for these areas and then verify (through test and audit activities) that they have been met. It is by influencing these design characteristics during the design phases of the project, that we can measure user and system performance, assess design improvements and influence the ultimate concern of effective task performance.

The emergent properties of the system (i.e. situation awareness, workload, teamwork and communication, and user confidence) cannot be directly designed, but are the product of a number of aspects of the system design and the design process. Human performance (both safe and effective) sits at the top of the framework and is NATS' ultimate concern. The NATS HF lead sets performance objectives and criteria for emergent properties at the beginning of the project to inform the design work. These performance objectives and criteria are then validated towards the end of the project.

Due to its complexity and impact, the iFACTS project spanned years, and resulted in a range of staff working on it, including different human factors specialists. The consistency of approach was a must and a pillar of quality assurance over the decade of project life-cycle. To enable the required consistency of approach, human factors work on the project was structured around identification, management and fulfillment of relevant requirements, all structured around impacted areas of the HF assurance framework.

Relevant requirements originated throughout the project life cycle from human factors work, user-focused workshops, training needs analyses, process and procedural changes identified, and a range of risk management workshops. As the project matured, the more specific the requirements became. In the final phases of the project, a complete set of requirements was available for system validation, managed via IBM DOORS application. The type and nature of requirements was driving the design of system validation scenarios. Some of the requirements required specific event to be simulated for validation evidence to be captured, whilst others were quite generic and able to be captured in each and every scenario. Each requirement was mapped with relevant measures coming from subjective, observational/behavioral, and objective classes of measures.

Figure 2. Comparison on system validation and live operational data for the three key HF measures



Once the relevant measures were selected, it was necessary to define a target range for each measure of interest. This allowed the human factors specialists to set success criteria for validation of each requirement. This was usually based upon the baseline captured using the existing system, fine-tuned based on expert judgment and known properties of traffic. The target for each measure of interest was always defined prior to the beginning of each system validation activity.

iFACTS human factors assurance status was based on evidence derived from several sources: R&D work, past research, user-centered design, safety work, operational documents, system validation, and monitoring of live operations. In the final stages of the project, system validation evidence became vital. Figure 2 shows an excellent transfer of human performance findings (Workload, Situational Awareness (SA) and User Confidence) from system validation of iFACTS to its live operations.

CONCLUSIONS

Unless the activities that human factors professionals engage in within ATC acquisitions from initial requirements, design through to operational implementation, result in acceptable levels of human/task performance, we are not being as effective as a community as we should. Successfully bringing together human factors specialists, project managers and acquisition specialists requires compromise and useful goals that support and not hinder large organizational acquisition programs.

The NATS approach is a pragmatic one, combining academic knowledge with real world compromise. Acceptable levels of task performance were defined by human factors specialists. For complex systems or tools, lower level of task performance can be set as acceptable for the "O" date with appropriate operational mitigations. The higher levels of task performance should be expected within a predefined time in operations, when the operational mitigations, previously put in place, can be eliminated. This allows an incremental and 'soft 'implementation of complex systems rather than a 'big bang'. A key component to their approach to the application of human factors is traceability of all requirements within a project, all the way to documentation and final testing results (including verification and validation) as well as throughout monitoring of the system in live operation.

Historically there have been challenges to measuring performance in ATC human factors and the human factors community has not settled on a standard way to test ATC systems nor a standard set of performance measures used to assess whether a system performs successfully.

Within the FAA, we need a consistent method for testing new systems and human performance measures to support whether the requirements have been fulfilled. We must concede that we will never find a method that everyone universally agrees with, disagreements will be based around robust academic arguments. However, to acquisitions managers, those arguments are just that, academic. Technology is moving forward, systems are being procured and the longer it takes us to agree on such methods, the more systems are being operationally installed, in ways that are not what human factors experts would consider optimal. We need to select a consistent way to test new systems and a valid way to evaluate them for success. A model of human factors in acquisitions that drives through the whole process, defining and supporting each stage of acquisitions with appropriate tools/defined activities, is essential.

References

Hopkin, V. D. (1980). The measurement of the air traffic controller. Human Factors, 22(5), 547-560.

Krokos, K. J., & Baker, D. P., Norris, D. G., & Smith, M. A. (2007). *Development of Performance Standards for Air Traffic Control Specialists* (FAA Technical report submitted under Grant # 99-G-048). Washington, DC: American Institutes for Research.

Acknowledgements

The authors would also like to point out that the views expressed in this paper are primarily the views of the authors and may not represent the views and opinions of the FAA, NATS or AIR.

THE EFFECTS OF STEREOSCOPIC RADAR DISPLAYS ON AIR TRAFFIC CONTROLLER PERFORMANCE

Jason G. Russi, TSgt, USAF; Brent T. Langhals, Lt Col, USAF; Michael E. Miller, PhD
Air Force Institute of Technology
Dayton, Ohio

Mr. Eric Heft
711th Human Performance Wing, WPAFB, USAF
Dayton, Ohio

Controllers identify vertical separation in aircraft depicted on 2-D radar displays by calculating altitude from numerical values. This is used to create a 3-D mental image to determine vertical spacing; a mentally fatiguing practice. Current stereoscopic display technology exists that may allow reduction of this aspect of controller workload. With a near doubling of traffic expected within the next two decades (FAA, 2012), controllers' abilities to rapidly interpret spacing and maintain awareness will become increasingly imperative to safety. A stereoscopic radar simulator was developed and field-tested with 35 USAF controllers. It presented a top-down view, similar to traditional radar displays, however, altitude was depicted through stereoscopic disparity, permitting vertical separation to be viewed, rather than calculated.

Air traffic growth is expected to increase 90% over the next two decades (FAA, 2012). This increase in traffic volume and the limited capacity of the current national airspace system (NAS) has driven the Federal Aviation Administration (FAA) to plan a comprehensive overhaul of Air Traffic Control (ATC) operations (Marsa, 2009). The FAA's Next Generation Air traffic Control or "NextGen" initiative is an effort designed to increase the NAS capabilities; especially with regards to safety and aircraft efficiency. New technology incorporated under NextGen will permit aircraft to continuously broadcast aircraft location, speed, and flight information and provide this information in each aircraft to create automatically self-separating traffic (Mc Callie, 2011). Therefore controllers are expected to transition to passive roles of monitoring self-separating traffic and become responsible for monitoring larger numbers of aircraft. As a result, the workload, tasks and workstations of these controllers will change dramatically (Prevot et al, 2008).

The impact of NextGen driven changes upon controller responsibilities and duty requirements must be clearly understood to ensure minimal impact to flight safety. Currently, air traffic controllers must rapidly interpret vast amounts of data and dynamic traffic information to actively direct air traffic away from detected conflicts. Under NextGen, despite automatic self-separating aircraft, controllers will still be required to monitor the same data, but will only issue directions if the system fails to initiate separation actions. Unfortunately, research has shown that human vigilance wanes rapidly when active engagement, in this case issuance of instructions, is omitted (Parasuraman et al, 1993).

It is therefore, imperative to develop improved methods for delivering information to air traffic controllers that mitigates the negative impacts of passive monitoring. Final responsibility for flight safety will always remain with human controllers; thus their ability to observe the increased traffic volume accurately and in an intuitive manner will be increasingly critical for maximizing efficiency and safety. Toward this end, a simulated stereoscopic radar workstation was developed and field-tested with 35 USAF controllers. This workstation presents a view similar to traditional radar displays but depicts altitude through the use of stereoscopic disparity, permitting vertical separation to be visually represented as plainly as lateral separation.

The goal of this study was to determine the potential impacts of this stereoscopic workstation. Of particular interest were impacts to fatigue reduction, decision-making, perceived workload and situation awareness. From a human factors (HF) perspective, the displays must not only be intuitive, they must also accurately present the information and raw data the controller needs to support rapid decision making. For air traffic controllers this equates to depicting aircraft to permit easy understanding and projection of lateral and vertical separation. Therefore the research question is: *Does stereoscopic presentation of digital radar displays enhance controller performance and effectiveness?*

Background

ATC is a dynamic environment where controllers constantly receive a large volume of information from multiple sources to monitor changes, make decisions, and perform effective actions (Xing & Manning, 2005). ATC systems must be designed to provide the operator with the information and capabilities necessary to support the controllers (Endsley & Garland, 2000). The controller must formulate an accurate representation of the situation, given many sources of information about aircraft, sectors, and flight rules and then direct aircraft to maintain a safe airspace. These activities are best described as the mental modeling that a controller must perform to comprehend the aircraft positions in spatial orientation to one another.

Stereoscopic Displays

Research has demonstrated that stereoscopic display technologies are considered, "indispensible" for viewing complex and extensive high-dimensional scientific data and objects especially in a dynamic and temporal (4-D) capacity (Chau et al, 2012). However the applicability of stereoscopic viewing for the purpose of air traffic control potential has been met with mixed results (Parker & Wallis, 1948; McIntire et al, 2012). In an ATC context side-view combined with a top-view, commonly referred to as "coplanar", slant view and pilot perspective stereoscopic methods have been evaluated. An inherent limitation to slant view or pilot perspective stereoscopic methods is that lateral separation loses scale as aircraft are depicted farther away just as in real life the targets become smaller and more difficult to discern at a distance. This limitation renders these views useless for the purpose of determining whether a minimum separation exists between aircraft. According to Tavanti et al, there is a consistent deficiency in this type of display with regards to the unknown distance along the depth axis affording little to no reference to the horizon line (2003).

Virtual reality (VR) caves and other types of semi-immersive control interfaces have also been investigated for air traffic control operations and have been shown to have limited applicability (Persiani & Liverani, 2000). These display methods, aside from being physically cumbersome and requiring complete body reposition for 360 degree observance, the displays are view limiting and often disorienting due to a lack of ground and focal reference. Furthermore, these immersive displays lack the ability to provide a large scale depiction containing an operator's entire area of responsibility within a single field of view (FOV).

If stereoscopic viewing is to be useful for air traffic control operations, it must simultaneously depict lateral and vertical separation. Recent advances in technology allow for improved "top-down" viewing of a 3D world. As a result, this study focuses on the effects of providing a top-down stereoscopic view for air traffic control operations. Such a system maintains the perspective available to controller's in today's system to allow the controller to rapidly grasp the lateral separation, while using stereoscopic disparity to depict altitude, permitting the controller to assess the vertical separation based on perception of depth afforded by the stereoscopic top-down view. The following sections highlight the expected effects of incorporating stereoscopic into traditional ATC displays through a top-down view

Situation Awareness

Endsley indicated that operators and decision makers are often bombarded with far too much data to sort, leaving them less informed as they are unable to quickly access the information they need in a timely manner (1995). Situation awareness (SA) is the focus of an air traffic controllers training and the development of this skill, as well as the ability to maintain it, is critical to airspace safety. SA is defined by Endsley as, "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." (2000). Creating a stereoscopic view of a traditional top-down digital radar displays affords the controller the best possible perspective, preserving both the scale and lateral separation while providing a scale-accurate depth perspective to assess vertical separation. By providing simultaneous vertical and lateral separation, it is hypothesized that situational awareness will be enhanced.

H1: The use of stereoscopic digital radar displays will have a positive effect on controller situation awareness and conflict detection.

Mental Modeling, Memory and Recall

Hopkin stated that a controller's ability to forget the last altitude or vector given to an aircraft permits them to replace previous data with the current or next assigned characteristics (Hopkin, 1980). This memory management may allow a controller to maintain better awareness of their assigned traffic (Gronlund et al, 1997). Rantanen and colleagues described a controller's cognitive task as a producing a temporal mental picture because temporal demands require controllers to "anticipate aircraft trajectories and pilot intentions well into the future, plan their actions, and then execute the planned actions at a proper time and in an appropriate sequence" (Ratenan et al. 2004).

This is commonly referred to as “The Picture” by air traffic controllers. Specifically, Shorrock states that there is an absence of emphasis on developing mental imagery and adapting it into technology design (2007). In traditional ATC, controllers build mental models, specifically altitude depiction, by complex mathematical calculations based upon text cues. Stereoscopic displays as investigated here eliminate the need for such calculation and will present the 3D mental model intuitively. Given the importance of the mental model to air traffic controller performance and the effect of recall accuracy on establishing a reliable mental model, hypothesis 2 is as follows:

H2: The use of stereoscopic digital radar displays will have a positive effect on controller recall accuracy of aircraft vertical position.

Perceptions of Workload and Task Difficulty

It has been extensively studied and agreed upon by many in the human performance field, that temporal and mental demands comprise the largest portion of a controller’s workload in ATC (Rantanen et al, 2004). The limiting factor to the growth of the air traffic volume is also known to be the air traffic controllers’ limitations in controlling aircraft in a given sector. It has been determined as recently as 2004 that the maximum safe operating number of active aircraft per operator is 15 at a time (Erzberger, 2004). Workload for controllers is often driven by the amount of mental calculations they must perform. An appropriate stereoscopic display reduced the necessary number of mathematical calculations, thus reducing overall controller workload. If workload is reduced, the optimal number of aircraft per controller may be increased. In order validate this belief, it is hypothesized that:

H3: The use of stereoscopic digital radar displays will decrease controller perceived workload.

H4: The use of stereoscopic digital radar displays will decrease controller perception of task difficulty.

Methodology

Participants

For this study the sample population was provided by the ATC technical training school at Keesler Air Force Base, MS. The study involved 35 USAF ATC instructors and students. Participation was voluntary and the study lasted one week. A primary advantage of this population was that all participants were familiar with standard air traffic procedures and displays. Experience levels varied from 2 months, to 30 years. The average years of ATC experience was 7.29 years. The average age of participants was 29.1 years and 26 of the 35 participants were male.

Experiment

The experiment employed a within subjects design with two conditions: control (standard display) and experiment (top down stereoscopic display). The experiment was balanced with half the participants assigned to either the control or experimental condition first to limit any learning or order bias. All scenarios involved a pre-recorded ATC scenario to limit the participants’ interaction (i.e. participants could not alter aircraft trajectory on screen). Each scenario lasted approximately eight minutes. Participants were provided with a brief training period prior to the start of the study. The experiment was designed to evaluate the controller recognition of vertical separation conflicts when viewing a traditional 2D top-down ATC display versus a top-down stereoscopic ATC display. A pilot study was conducted with six additional volunteers to ensure validity and provide feedback to permit the display design to be refined.

Equipment

Commercial-off-the-shelf equipment was used with the exception of the SIGNAL FAA simulation software, which is was provided by the FAA. The remainder of the suite included an ASUS VG236 23.5",120Hz stereoscopic LCD display monitor, and NVIDIA 3D Vision software, USB Controller, IR Emitter and associated active shutter stereoscopic glasses operated from a Dell Precision M4600 mobile workstation.

The experimental scenario was provided stereoscopic images with active shutter 3-D glasses on the ASUS 23.5” monitor. This displayed provided a top-down display with a ground reference plane depicted by range rings having near zero disparity and the aircraft having disparity that is proportion to altitude such that the aircraft “protruded” from the display above the ground plane. The aircraft were provided with accompanying data blocks, which were assigned the same disparity as the aircraft. Through this depiction, the lateral and vertical separation of the aircraft were equally evident without the need to read and calculate numerical values to determine altitude separation.

Figure 1. **The experiment equipment in use.**

Data Collection

Data was collected with pre- and post-treatment questionnaires. The pre-test questionnaire collected participant demographics as well as relevant and potentially influential variables. In addition to questionnaires, single mouse clicks, used by participants to indicate when they detected potential conflicts, were the only real-time measurements. After each scenario a questionnaire was administered to evaluate effectiveness of the technology through collection of specific performance measures. This questionnaire included a primary recall accuracy instrument, which was a screen capture from two minutes prior to the scenario end with the three-digit altitude codes blanked out, to be filled in from memory and questions regarding the relative locations of aircraft within the environment. The questionnaire also contained questions about the participants' perceptions of the technology and, confidence in their answers and a place for open ended feedback. There were no time limitations placed on any questionnaires.

Results

SPSS was used for all statistical analysis. The independent variable was display condition and dependent variables are indicated in each hypothesis. For all hypotheses, a *t*-test was first conducted with a 95% confidence interval (CI) to test for significant difference between control and experimental group means. If significance was found, a stepwise regression was performed to determine the influence of participant demographic and personality measures on the dependent variable. Distribution normality was assumed as the t-test is considered robust with respect to the assumption of normality, and homogeneity of variance was evaluated as assessed by Levene's Test for Equality of Variances.

To investigate the hypothesis that "The use of stereoscopic digital radar displays will have a positive effect on controller situation awareness and conflict detection" it was predicted that the stereoscopic display would increase the accuracy of the detection of on-screen conflicts. As the paths of the aircraft within the scenarios were depicted to provide a pair of lateral position conflicts, but never to provide conflict in both lateral position and altitude, the operators should not have had to indicate a conflict. Therefore, a higher number of clicks indicated a lower awareness of actual conflicts. The t-test indicated, the number of clicks was larger for the control group, 3.40 clicks with a standard deviation of 1.33, than the experimental group, mean of .91 clicks with a standard deviation of 1.27 ($t(68) = 7.991$, $p < .01$). This finding indicates a significant difference in SA and conflict detection performance between the display conditions. The stepwise regression model resulted in adjusted R^2 of .613 with Fatigue, Multitask, and Years of Experience in ATC included in the model at ($F(69) = 28.357$, $p < .01$).

To test the hypothesis that "The use of stereoscopic 3-D digital radar displays will have a positive effect on controller recall accuracy of aircraft position", we investigated the effect of display condition on recall accuracy. Surprisingly, the t-test indicated the control group answered a higher number of questions accurately, recalling and average of 4.81 questions with a standard error of 1.11, as compared to the stereoscopic condition, which averaged 2.73 correct questions with a standard error of 1.15 ($t(68) = 7.679$, $p < .01$). The stepwise regression resulted in an adjusted R^2 of .562 with depth perception deficiency and SA/distraction included as significant factors in the model at ($F(69) = 30.489$, $p < .01$).

Several participants reported having trouble recalling the exact location of the highest or lowest aircraft in the scenarios after viewing the stereoscopic display condition. They attributed the obvious safe vertical aircraft separation to their inability to recall. Participants in the stereoscopic display condition reported immediately recognizing aircraft that were operating in safe vertical proximity and dedicated fewer attentional resources to tracking and monitoring those targets, thus reducing memory recall of those tracks. Since it has been claimed that forgetting information may be just as important as remembering it in a dynamic memory situation like ATC (Hopkin, 1980), as the status of each aircraft track changes so rapidly, recalling the last or last several altitudes, (or other flight characteristics), may interfere with the controller's ability to remember the current or most recent altitude. It is therefore possible that the stereoscopic display condition allows the controllers to rapidly determine priorities, or "filter through the chaff", based upon which aircraft are more likely to require separation instructions from those which are clearly operating with safe separation. However, this finding requires further investigation.

To test the hypothesis that the use of the stereoscopic digital radar display would reduce perceived workload we investigated participants' post-conditions assessment of perceived workload. The control group reported a higher perceived workload, average rating of 4.74 with a standard error of 1.15, than the experimental group, 2.03 with a standard error of 1.89 ($t(68) = 8.61$, $p < .01$). The stepwise-regression included self-reported distraction along with display condition as significant and showed an adjusted R^2 of .598, which was significant ($F(69) = 52.236$, $p < .01$). Controllers reported that the experimental treatment "seemed" easier and they were able to comprehend vertical separation more readily, making the tasks seem simpler.

To test the hypothesis that the stereoscopic digital radar display will reduce the perception of task difficulty, the rating of task difficulty was evaluated. It was found that the control group reported a higher level of task difficulty, reporting a mean value of 5.31, with a standard error of 1.79, than the experimental group, with a mean value of 3.61 and a standard error of 1.00 ($t(53.271) = 4.894$, $p < .01$). A stepwise-regression resulted in a model which included self reported distraction, Aviation Experience, and Computer Use to produce an adjusted R^2 of 360 ($F(69) = 10.695$, $p < .01$).

It is noteworthy that not only were the majority of the hypotheses confirmed during this experiment, but the participants valued the innovation. When asked if they found the stereoscopic display useful in the performance of their task, 88.6% of the air traffic controllers answered affirmably.

Conclusions

This was an exploratory study showing strong potential of top-down view, stereoscopic displays to increase controller performance when performing vertical and lateral separation comprehension and conflict detection. The reduced perceived workload and task difficulty indicate that the use of such a display may have the potential to permit controllers to monitor more aircraft with potentially less loss of awareness. Furthermore the study also implies that controllers, with the help of a more intuitive and less mentally tasking display, may be better prepared to maintain vigilance over larger sectors for longer periods of time. Finally from a safety standpoint the increased awareness and detection of potential conflicts also indicate that controllers may be better equipped to maintain safe separation of aircraft thereby permitting an increase in traffic safety in the system as a whole. However, these implications require further study with more realistic scenarios.

Acknowledgements

The authors would like to thank the Air Force Air Traffic Control community for supporting this research. Without their support for the topic and their willingness to provide access to the participants at Kessler AFB, this research would not have been possible. Additionally, the authors would like to the 711th Human Performance Wing for their contributions of equipment, lab space, and expertise. Finally, we would like to acknowledge the Federal Aviation Administration for providing the SIGNAL training software upon which the experiment was based.

References

Chau, D., McGinnis, B., Talandis, J., Leigh, J., Peterka, T., Knoll, A., Jellinek, J. (2012). A simultaneous 2D/3D autostereo workstation. Paper presented at the *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, , 8288* 89.

Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. *Situation Awareness Analysis and Measurement,* 3-32.

Erzberger, H. (2004). Transforming the NAS: The next generation air traffic control system. Paper presented at the *Proceedings of the 24th Int. Congress of the Aeronautical Sciences (ICAS),*

Federal Aviation Administration. (2012). *FAA aerospace forecast fiscal years 2012-2032* (1st ed.). Washington, D.C.: FAA.

Gronlund, S. D., United States. Office of Aviation Medicine, Civil Aeromedical Institute, & University of Oklahoma. Dept. of Psychology. (1997). *The role of memory in air traffic control*. Washington, D.C.; Springfield, Va.: U.S. Dept. of Transportation, Federal Aviation Administration, Office of Aviation Medicine; Available to the public through the National Technical Information Service. Retrieved from http://purl.access.gpo.gov/GPO/LPS84902

Hopkin, V. D. (1980). The measurement of the air traffic controller. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 22*(5), 547-560. doi: 10.1177/001872088002200504

Marsa, L. (2009). Wing and prayer. *Discover, 30*(8), 58-65. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=43977857&site=ehost-live

McCallie, D. L. (2011). *Exploring potential ADS-B vulnerabilities in the FAA's NextGen air trasnportation system.* (Unpublished MASTERS). Air Force Institute of Technology, Dayton, OH.

McIntire, J. P., Havig, P. R., & Geiselman, E. E. (2012). What is 3D good for? A review of human performance on stereoscopic 3D displays. Paper presented at the *Proc. of SPIE Vol, , 8383* 83830X-1.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced'complacency'. *The International Journal of Aviation Psychology, 3*(1), 1-23.

Parker, E., & Wallis, P. (1948). Three-dimensional cathode-ray tube displays. *Electrical Engineers-Part III: Radio and Communication Engineering, Journal of the Institution of, 95*(37), 371-387.

Persiani, F., & Liverani, A. (2000). Semi-immersive synthetic environment for cooperative air traffic control. Paper presented at the *Proceedings of 22nd International Congress of Aeronautical Sciences, Harrogate, UK, , 27* 08-01.

Prevot, T., Homola, J., & Mercer, J. (2008). Human-in-the-loop evaluation of ground-based automated separation assurance for NEXTGEN. Paper presented at the *Congress of International Council of the Aeronautical Sciences Anchorage, Anchorage, AK,*

Rantanen, E. M., McCarley, J. S., & Xu, X. (2004). Time delays in air traffic control communication loop: Effect on controller performanceand workload. *International Journal of Aviation Psychology, 14*(4), 369-394. doi: 10.1207/s15327108ijap1404_3

Shorrock, S. T., (2007). Errors of perception in air traffic control. *Safety Science, 45*(8), 890-904. doi: 10.1016/j.ssci.2006.08.018

DESIGN AND EVALUATION OF A CO-PLANAR SEPARATION DISPLAY

J. Ellerbroek, K. C. R. Brantegem, M. M. van Paassen, and M. Mulder
Faculty of Aerospace Engineering, Delft University of Technology,
Kluyverweg 1, 2629 HS, Delft, The Netherlands

This paper describes the design and evaluation of a co-planar constraint-based airborne separation assistance display. The display is a combination of previous single-plane presentations, with additional visualization of the interactions that exist between these planes. Each of these displays combines a spatial representation of the airspace with a velocity maneuver space, that relates own aircraft maneuver variables to the shape and affordances of the airspace. The evaluation presented in this paper consisted of two experiments: an active conflict resolution task, and a passive SA assessment. Both experiments compare the co-planar concept with a baseline display that is very similar, but does not visualize planar interactions. Results showed that although pilots performed well with either display, performance was consistently better with the augmented display.

The work presented in this paper is part of an ongoing study, which applies a constraint-based approach to the design of a 3-D airborne separation assistance interface. In this paper, a co-planar display is proposed that presents constraints on maneuvering in a `velocity action space', that is overlaid on traditional moving-map displays. This paper also presents the results from an evaluation of the display concept.

In response to continuously growing levels of automation, several studies argue that proper human-automation interaction, and appropriate trust in automation require transparent and understandable functioning of automation. The interface should provide a window to the reasoning and functioning of the automation (Inagaki, 2006). The constraint-based displays proposed in this study aim to improve pilots' understanding of automated resolutions, by helping them understand how different elements in the work environment interact, and shape the possibilities for conflict resolution. The evaluation of such a display should therefore focus on how elements of the display affect the operator's understanding of the traffic situation. The experiments in this paper were designed to serve this purpose. An active conflict resolution experiment was performed to evaluate how operator performance and behavior are influenced by the visualization. A second experiment consisted of a passive situation awareness assessment (See also Ellerbroek et al. (2013)).

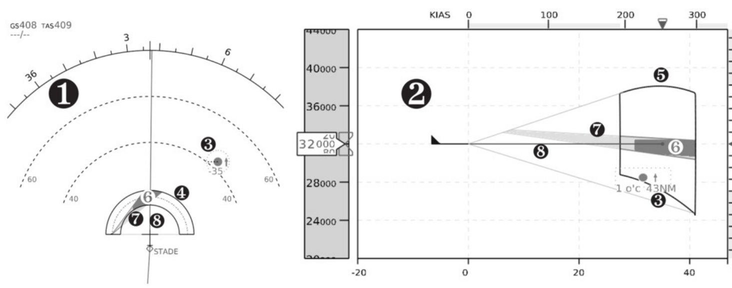


Figure 1: Concept for a co-planar separation assistance display. This figure shows a Horizontal Situation Display (❶), and a Vertical Situation Display (❷), with added separation assistance overlays (❹ and ❺).

The Interface

Fig. 1 illustrates a concept for a self-separation interface, that presents separation-related constraints and relations on a co-planar display. Important elements of the display are numbered in the figure, and are briefly described in this section. See Ellerbroek et al. (2013) for a more elaborate review of this display. In this concept, the 3-D traffic situation is visualized in two orthogonal, 2-D views: a top-down view (❶), and a side view (❷). Both views present a classical ownship-centered moving map, that shows spatial information such as the planned route and the relative positions of other aircraft (❸). In addition, maneuvering constraints are shown on both displays through velocity action-space overlays (❹, ❺), referred to as State-Vector Envelopes (SVEs) in the remainder of this text.

The horizontal SVE (❹) shows the horizontal maneuver space, in terms of track angle and airspeed. Its boundaries are determined by the aircraft performance limits: The aircraft minimum and maximum operating speeds result in the concentric circular boundaries of the SVE. The vertical SVE (❺) shows a vertical maneuvering space, in terms of airspeed and vertical speed. Similar to the horizontal SVE, its boundaries are also determined by aircraft performance limits. The vertical edges of the SVE result from the limits on aircraft airspeed. The curved edge at the top of the vertical SVE visualizes the maximum steady climb at each velocity. The bottom edge indicates maximum steady descent. Combined, the areas within these envelopes describe all reachable velocity vectors of the aircraft.

Intruder aircraft that are within detection range will reduce the available maneuver space in the horizontal and vertical SVEs. The reduced forbidden areas (RFAs) (❻) give the most precise representation of these constraints. On the Horizontal Situation Display (HSD), an RFA gives the constraints imposed by an intruder on ownship track angle and airspeed, for the current value of ownship vertical speed. On the Vertical Situation Display (VSD), an RFA gives intruder-imposed constraints on ownship airspeed and vertical speed, for the current ownship heading. These RFAs result from the intersection between a flat cutting-plane, and the 3-D forbidden area (FA): the set of ownship velocities that result in a conflict with the corresponding intruder, see Figures 2 and 3. These RFAs are subsets of the projected FAs, that were used in the earlier constraint-based separation displays (Dam et al. 2008).

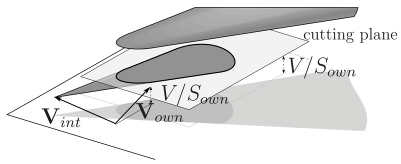


Figure 2: Horizontal constraints using a cutting-plane.

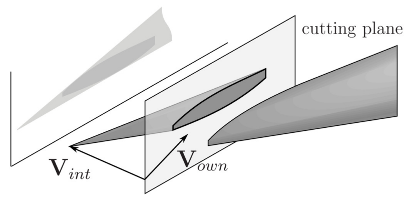


Figure 3: Vertical constraints using a cutting-plane.

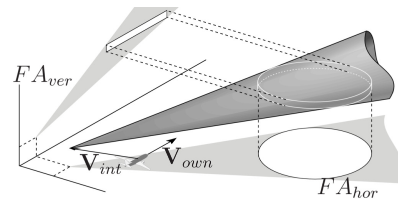


Figure 4: Horizontal and vertical projections of traffic constraints.

The projected FAs (❼) are still shown in combination with the RFAs, as they provide several Situation Awareness (SA) related cues, as well as an outer limit on the shape and size of the RFA, when a perpendicular flight parameter is modified. Figure 4 shows how these projections are constructed.

Conflict urgency is explicitly indicated on the display using intruder symbology similar to the existing Traffic Collision Avoidance (TCAS) system. In addition, conflict urgency is also indicated using color coding for all of the display elements that correspond to one intruder. This means that the aircraft symbols on both displays, as well as the FA triangles and RFAs on both displays are colored according to the urgency of the conflict between ownship and the corresponding intruder.

Experimental Evaluation

Two experiments, an active conflict resolution task and a passive situation awareness assessment, were conducted to evaluate the co-planar separation assistance display. Both experiments compared the new concept to a baseline display that was very similar, but did not show how constraints interacted between projection planes.

Method

Both the active conflict resolution experiment and the passive SA assessment were designed as a within-subjects repeated-measures, where factors *display type* and *conflict geometry* were varied. In both experiments, *display type* had two levels: reduced forbidden areas could be either present or absent. This factor was introduced to illustrate the effect of the additions that the co-planar concept features compared to the original 2-D separation displays. In the active experiment, the *conflict geometry* factor differentiated between *simple* and *difficult* scenarios, and between phases of flight (climb, cruise, and descent), resulting in 6 scenarios (2 X 3). Simple conflicts always featured only one intruder, close to the own flight level and track. In difficult scenarios, three intruders were present in each scenario, which could be both off-level and off-track. In the passive SA assessment, conflicts could be either on-track or off-track, and on-level or off-level, resulting in 4 scenarios (2 X 2).

In both experiments, conditions were presented in a randomized block design. In the active experiment, trials were combined in four blocks of four sequential conflict scenarios. Each block started with a climb from flight level FL220 to flight level FL320, at 1,000 ft/min, followed by a cruise segment, and then a descent back to flight level FL220, again at 1,000 ft/min. Each segment featured at least one conflict. A block lasted approximately 40 minutes. The *display type* factor was kept constant over two blocks: first two blocks with one display, then two blocks with the other. The order of presentation for the display types was varied evenly over the subjects.

In all scenarios in both experiments, multiple options, both horizontal and vertical, were available to solve the conflict, although not all options were equally fast and efficient. In the active experiment, intruder aircraft never maneuvered in order to solve a conflict situation, instead they just kept following their initial path.

Dependent measures. *Resolution strategy* was an objective measure, measured in terms of velocity vector change dimensions. This could be a combination of a change in heading, speed and vertical speed (V/S). Path deviation and the initial reaction time were used as measures of *performance*. The path deviation metric differentiates between horizontal and vertical maneuvers: For horizontal maneuvers, the path deviation was characterized by the additional distance flown. In case of a vertical maneuver during the climb or descent phase, the mean deviation from the prescribed V/S was used. For cruise conflicts, the maximum altitude deviation from the cruising level was measured. Pilot reaction time and the total time of the resolution maneuver, i.e., the time between leaving and rejoining the reference trajectory, were used as metrics that allow for comparison between vertical and horizontal maneuvers. *Safety* was measured in terms of minimum separation, and the occurrence of losses of separation.

SA questions in the passive assessment relate to easily identifiable information such as relative intruder position and intruder velocity, but some questions also required the subject to use information cues to predict the outcome given the current situation. The questions were categorized using Endsley's levels of SA (Endsley, 1995), and the subject's certainty of his answer was recorded together with the answers. The combination results in a grade, that categorizes answers into four groups, following Hunt's method of measuring knowledge (Hunt, 2003).

Hypotheses. Several previous studies found that pilots prefer single-axis maneuvers, keeping velocity constant (Alexander et al., 2005; Ellerbroek et al., 2011). It was therefore hypothesized that most maneuvers would be either heading-only, or V/S-only. It was also hypothesized that maneuver choice would depend on phase of flight, i.e., that climb and descent conflicts would be solved vertically and that cruise conflicts would be solved horizontally.

Differences between displays were only expected during difficult scenarios. It was therefore hypothesized that performance would be improved with the augmented display in difficult scenarios, and that SA would be higher, especially at the projection level. Because the RFAs show more precise constraints than the projected FAs, it was also hypothesized that they would result in lower separation at the Closest Point of Approach (CPA), as previous studies showed that the precision with which constraints are presented is used by pilots to optimize their efficiency (Ellerbroek et al., 2011). The number of separation losses was hypothesized to be low, regardless of display.

Because SA level 1 questions relate to elements that are directly perceivable on both displays, it was hypothesized that SA level 1 score would be high, regardless of display type. Since the augmented display visualizes more higher-level information and relationships, it was also hypothesized that the SA scores between displays would diverge increasingly, with higher SA levels. An interaction with scenario was expected for this effect, as the difference between displays becomes increasingly pronounced for scenarios with off-level or off-track intruders.

Results

Kolmogorov-Smirnov tests on the ratio data revealed that a normality assumption could never be made (altitude deviations, CPA values, response times and resolution times, $p < 0.001$ in each case). Therefore, only non-parametric tests were used: the Wilcoxon Signed-Rank test (test statistic z) for metrics based on ratio data that did not depend on the chosen maneuver, and the Wilcoxon rank sum test (test statistic W) for all other metrics based on ratio data. Pearson's chi squared test (test statistic $\chi 2$) was used for categorical metrics. Effects were considered significant at a probability level $p <= 0.05$, where p is the probability that the null hypothesis is true.

Resolution strategy. Figures 5 and 6 show resolution strategy divided into five levels: *vertical maneuvers (with and without speed)*, *horizontal maneuvers (with and without speed)*, and *combined three-way maneuvers*. Maneuver selection will depend on conflict geometry, phase of flight, performance limitations, and personal or airline preference. Fig. 5 shows the maneuver choice for the simple cruise, climb and descent scenarios. The majority of the maneuvers for the climb and descent scenarios were V/S-only, regardless of display type (82% - 94%). With one exception, the direction of the change in V/S was always the same: the climb conflict was always solved by increasing the rate of climb, and the descent conflict by decreasing the rate of descent. These choices correspond to the smallest available state change for the current conflict, an efficiency strategy given to the subjects during the briefing. They can, however, also be an indication of a preference for `staying high', to optimize for fuel efficiency.

Although the spread in solution strategy was larger than in the climb and descent scenarios, the majority of the resolutions in the simple cruise scenario was still heading only (baseline display 53%, augmented 65%). As was hypothesized, phase of flight was an important factor when deciding on a solution strategy. Comparison between the cruise scenario and the vertical scenarios showed a significant difference in resolution decisions ($\chi 2(2) = 56.9$, p < 0.001). Comparison between displays did not reveal significant effects for simple conflicts.

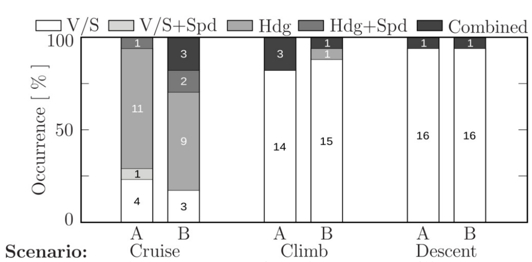


Figure 5: Solution strategy for simple conflicts, sorted by scenario and display type (A=augmented, B=baseline).

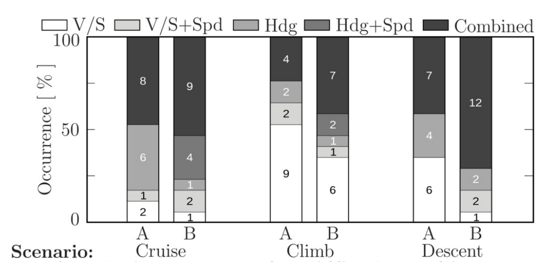


Figure 6: Solution strategy for difficult conflicts.

Fig. 6 shows the maneuver choice for the difficult cruise, climb and descent scenarios. In terms of resolution strategy, the difference between the displays is visible in the number of multi-axis resolutions (V/S+SPD, HDG+SPD, or combined), which were used significantly more often with the baseline display: 77% for the baseline display, compared to 43% for the augmented display, for the climb, cruise, and descent scenario combined ($\chi2(1) = 11.8$, $p = 0.001$). Most of these multi-axis resolutions were sequential maneuvers, rather than a single combined maneuver, regardless of display type. In other words, pilots often changed their minds after an initial resolution. The high amount of multi-axis resolutions, therefore, doesn't necessarily refute the hypothesis of single-axis maneuver preference, as the initial resolution maneuver often was single-axis. It is more likely that lack of training plays a large role in this result. The difference between displays in the number of multi-axis resolutions can also be indicative of reduced SA with the baseline display.

Based on pilot comments during the experiment, the multi-axis maneuvers can be classified into two categories. For the baseline display, the most often heard comment was that a pilot realized that he had made a wrong initial maneuver. This was either a maneuver that didn't resolve the conflict, or a maneuver that resulted in a very inefficient resolution. A second category of maneuvers were from pilots that attempted to increase efficiency, by maneuvering in an additional direction. Phase of flight also significantly influenced maneuver strategy in the difficult scenarios ($\chi2(2) = 6.3$, $p = 0.04$). The cruise conflict was solved horizontally (32.4%) almost twice as much as vertically (17.6%). Similarly, the climb and descent scenarios were more often solved vertically (39.7%) than horizontally (16.2%).

Safety. Fig. 7 shows a cumulative distribution graph of the normalized CPA values, for both displays. Separation was violated in 8 out of 272 measured trials, twice with the baseline display, and 6 times with the augmented display. In all eight cases, this occurred during a premature return to the nominal track, and in all cases, the incursion was minimal (all within 10% of the separation minimum, and 6 less than 1%). A common practice that was observed in this, but also in previous experiments with a constraint-based display (Ellerbroek et al., 2011), was that after resolving a conflict, pilots are inclined to optimize their performance by returning to their nominal state as soon as possible, in small steps, while staying as close as possible to the edge of the FA. In these situations, a judgment error can easily result in a (small) separation violation. The difference between displays in the number of losses of separation was not significant ($\chi2(1) = 2.1$, $p = 0.15$), but does illustrate that the more restrictive constraints on the baseline display can act as an added safety margin.

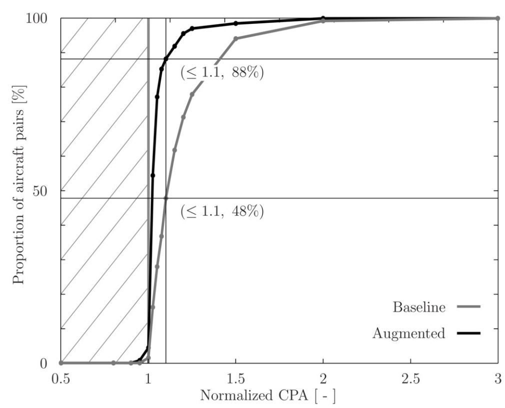


Figure 7: Cumulative normalized CPA.

Performance. Fig. 7 also shows that, especially with the augmented display, pilots often came within close distance of the protected zone of the other aircraft. With the augmented display, 88% came closer than 1.1 times the separation minimum, versus 48% for the baseline display. In terms of performance, this is a strong indication that pilots use the precise visualization of constraints to optimize the efficiency of their resolution. The difference in CPA distance between displays was significant ($z = -7.22$, $p < 0.001$). For horizontal maneuvers, path deviation never revealed a significant effect. The difficult descent and climb scenarios did show a consistent improvement of the augmented display over the baseline display, but contained too few samples to provide sufficient statistical power.

As climb and descent scenarios were mostly solved with a change in V/S, the mean deviation from the prescribed V/S will be used to observe differences in performance between displays for vertical conflicts. Although there is a consistent trend of the augmented display performing better than the baseline display, this difference was only significant in the difficult descent scenario ($W = 24$, $p = 0.024$). Cruise conflicts were solved 14 times out of 68 with a change in V/S. Although the mean deviation from the prescribed V/S did not reveal a significant difference, the maximum altitude deviation did differ significantly between display types, where the altitude deviation was always smaller with the augmented display ($W = 62$, $p = 0.029$). This is also an indication that pilots exploit the precise constraint visualization to optimize maneuver efficiency (Ellerbroek et al., 2011).

Reaction time and resolution duration are measures that can be considered independent of the maneuver dimension, and can therefore be used as overall metrics to compare display conditions. Here, resolution duration is a measure of performance of a maneuver, and reaction time is indicative of the difficulty experienced by pilots. As hypothesized, both these measures show significant effects of display type for the difficult conflict scenarios, but not for the simple conflict scenarios. For the simple conflict geometries, the two display variants show comparable maneuver constraints. It is therefore not expected that difficulty and resolution performance vary significantly between display types. For difficult scenarios, results for the augmented display show significantly shorter reaction times ($z = -2.32$, $p = 0.021$), and significantly shorter resolution durations ($z = -2.53$, $p = 0.012$).

Situation awareness. The SA scores from the experiment were grouped using Endsley's three levels of awareness (Endsley 1995), and are shown in Fig. 8, for each combination of display type and scenario. These SA scores will depend on conflict geometry and accuracy of the visualization, but also on other factors that influence the buildup of SA, such as attention and workload. As hypothesized, the first column in Fig. 8 shows that the majority of the subjects (92 - 100%) managed to achieve the highest SA score for level one questions, regardless of scenario or display. A comparison between displays for SA level 1 therefore also did not reveal significant effects, see Table 1.

A main effects analysis (see Table 1) showed that, as hypothesized, display becomes a significant factor for SA scores at levels two and three: As can be seen in Fig. 8, subjects scored consistently lower with the baseline display. A post-hoc analysis revealed that this effect increases when scenarios become increasingly off-level and off-track: Table 1 shows that the effect of display is only significant for level 2/3 scores in the off-level and off-track scenario. This supports the hypothesis that conflict geometry would influence SA scores between displays.

Table 1: Comparison between display types of the SA scores.

Level × scenario	**SA level 1**	**SA level 2**	**SA level 3**
Main effect	$\chi^2(1) = 0.4$ $p = 0.540$ ○	$\chi^2(1) = 10.7$ $p = 0.001$ ★★	$\chi^2(1) = 20.7$ $p < 0.001$ ★★
On-level/on-track	$z = -0.378$ $p = 0.705$ ○	$z = -0.556$ $p = 0.579$ ○	$z = -1.633$ $p = 0.102$ ○
On-level/off-track	$z = -1.000$ $p = 0.317$ ○	$z = -1.016$ $p = 0.309$ ○	$z = -1.173$ $p = 0.241$ ○
Off-level/on-track	$z = -1.000$ $p = 0.317$ ○	$z = -1.885$ $p = 0.059$ ○	$z = -2.362$ $p = 0.018$ ★
Off-level/off-track	$z = -0.136$ $p = 0.892$ ○	$z = -3.430$ $p < 0.001$ ★★	$z = -3.084$ $p = 0.002$ ★★

★★ significant; ★ marginally significant; ○ not significant

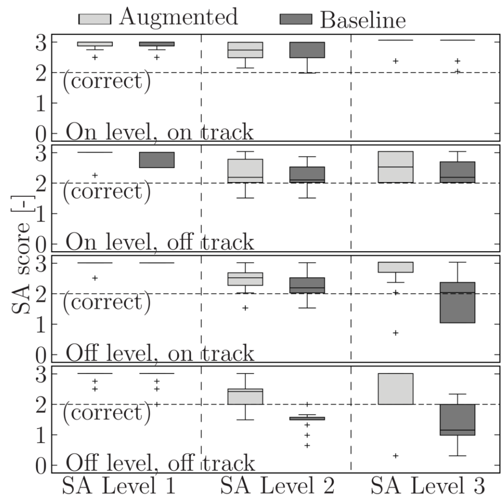


Figure 8: SA scores, averaged per pilot.

Discussion

The displays in this study are designed to help a pilot understand the reasoning behind automated decisions, by showing constraints and relationships within the work domain. This work domain information invariably forms the premise on which automation bases its actions, and is therefore also invaluable to pilots when they need to judge the automation's functioning. Although the experiment did not feature automated conflict resolution, and can therefore not be used to evaluate interaction between human and automation, the pilots' resolution decisions do give insight in how the information on the display is used by pilots, and how it affects their SA.

The objective measures presented in this paper show several trends. An effect that is seen in several other studies was that many resolution maneuvers were single-axis. Current results showed, however, that this effect diminished for more difficult scenarios. It can be argued that this was mostly a training issue, as pilot comments during the experiment often indicated that an erroneous initial resolution choice was made. Several pilots also mentioned in the post-experiment questionnaire that more training would be required to be able to understand and properly use the interface. Occasionally, pilots also initiated a multi-axis maneuver `just to see what happens', which can be considered an artifact of volunteer test subjects. In some cases pilots indicated that they made a multi-axis maneuver to improve efficiency. Path deviation measurements, however, showed that this was never the result.

Although difficult scenarios resulted in more multi-axis maneuvers, this effect did depend on display configuration, where multi-axis maneuvers were made more often with the baseline display. Since many of the multi-axis maneuvers were corrections of an erroneous initial single-axis maneuver, this can be an indication that, with the same (limited) level of training, pilots performed better with the augmented display. They made fewer errors, indicating a beneficial effect on traffic awareness of the augmented display.

As hypothesized, phase of flight had a significant effect on resolution choice, regardless of scenario difficulty. This preference can be seen as the result of a procedural constraint (i.e., phase of flight) that is however not directly visible on the display. This indicates that pilots can use the presented constraints, and apply them to other rules and procedures. This is classified by Rasmussen as Rule Based Behavior (Rasmussen, 1983). Ideally, the interface should support pilots at all levels of cognitive behavior, while not forcing them to control at a higher level than necessary (Vicente & Rasmussen, 1992).

A persistent result found in this experiment, and earlier experiments with a constraint-based display, is that after reaching a conflict-free state, the majority of the subjects returned to their original track in several small steps, following the edge of the constraint area as closely as possible (Borst et al., 2010; Ellerbroek et al., 2011). This behavior can be attributed to showing precise constraints: when maneuver limits are visualized with high precision, human operators will use that precision to maximize their efficiency. As a result, the majority of the CPA's stay within 110% of the separation margin (augmented 88%, baseline 48%). This `hunting' behavior, however, also gives rise to judgment errors, and consequently also losses of separation, which occurred 8 times in the experiment. Although the incursions were very small, this is still an undesired side effect of showing precise constraints.

The SA assessment revealed that display becomes a significant factor in complex scenarios, for high-level SA probes. These scenarios consist of off-track and off-level geometry, which reveal the difference between the projected FAs and the RFAs. In these situations, even though both displays present the same *type* of information (horizontal and vertical maneuvering constraints), they differ in the *accuracy* of that information. The information that is hidden in the baseline display can still be derived to some extent, but this requires additional cognitive work.

In comparison with the baseline display, the augmented display reveals more properties and relations that are inherent to the work-domain. In the search for a display that properly supports pilots' SA, the trade-off will always be between showing more information on the one hand, and maintaining a clear, understandable and uncluttered display on the other hand. The results in this study show that performance and SA benefit from the improved accuracy of the constraint visualizations, and that pilot behavior is consistent with previous evaluations of constraint-based displays. Together with the preference ratings from the post-experiment questionnaire, these results also give no indication that this increased accuracy forms a problem in terms of display clutter. Nevertheless, future design iterations should continue to focus on the trade-off between information density and clutter.

Acknowledgements

This work has been co-financed by the European Organization for the Safety of Air Navigation (EUROCONTROL) under its research grant scheme of the innovative studies programme. Additional funding for this project is provided by the National Aerospace Laboratory (NLR), The Netherlands. The content of the work does not necessarily reflect the official position of EUROCONTROL nor NLR on the matter.

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. *The International Journal of Aviation Psychology*, 15, 1–21.

Borst, C., Mulder, M., & Paassen, M.M. van. (2010). Design and Simulator Evaluation of an Ecological Synthetic Vision Display. *Journal of Guidance, Control and Dynamics*, 33(5), 1577–1591.

Dam, S.B.J. van, Mulder, M., & Paassen, M.M. van. (2008). Ecological Interface Design of a Tactical Airborne Separation Assistance Tool. *IEEE Transactions on Systems, Man, and Cybernetics, part A: Systems and Humans*, 38(6), 1221–1233.

Ellerbroek, J., Brantegem, K.C.R., Paassen, M.M. van, Gelder, N. de, & Mulder, M. (2013). Experimental Evaluation of a Co-planar Airborne Separation Display. *IEEE Transactions on Human-Machine Systems*, 1(3).

Ellerbroek, J., Brantegem, K.C.R., Paassen, M.M. van, & Mulder, M. (2013). Design of a Co-Planar Airborne Separation Display. *IEEE Transactions on Human-Machine Systems*, 1(3).

Ellerbroek, J., Mulder, M., & Paassen, M.M. van. (2011,May). Evaluation of a Separation Assistance Display in a Multi-Actor Experiment. In *Proceedings of the 16th international symposium on aviation psychology*.

Endsley, M.R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37(1), 32–64.

Hunt, D.P. (2003). The Concept of Knowledge and How to Measure It. *Journal of Intellectual Capital*, 4(1),100–113.

Inagaki, T. (2006). Design of Human–Machine Interactions in Light of Domain-Dependence of Human-Centered Automation. *Cognition, Technology & Work*, 8(3), 161–167.

Rasmussen, J. (1983). Skills, Rules, Knowledge; Signals, Signs, Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13, 257–266.

Vicente, K.J., & Rasmussen, J. (1992). Ecological Interface Design: Theoretical Foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(4), 589–606.

USING ANIMATED GRAPHICS ON AIRCRAFT NAVIGATION DISPLAYS; PROS AND CONS

Chang-Geun Oh
Wright State University,
Dayton, Ohio 45435

An advantage of animations is improving an operator' ability to reconstruct information space. However, research also shows there can be disadvantages to the use of animation including the mismatch of the operations concepts of speed or motion of the task and the perceptions of the animation. An exploratory study was conducted to determine pilot opinions of whether animation would help them to respond to a route change instructions from Air Traffic Control rapidly and with high levels of interpretation accuracy. Many pilots indicated that animation would support their performance, but that animation should be optional and pilots should be able to control the speed of the animation.

Pilots communicate with air traffic control (ATC) through datalink communications (DataComm) as well as voice communications, and must perform successful interpretation of navigation clearances for the safe flight. Aircraft depict the current flight plan to pilots via their navigation displays (NDs). When pilots receive navigation clearances, pilots use the ND to support their interpretation of the clearance compared to the current flight plan. If pilots receive complex route clearances, they must perform complicated spatial reasoning. This may impose considerable pilot workload and misinterpretation of the clearance is possible (Gallimore, Kiss, Munoz, Oh, Crory, Ward, Green, Shingledecker, & Tsang, 2013). Even delayed interpretation can cause renegotiation with ATC. To support pilot clearance interpretation, novel forms of complex routing presentations need to be explored. Animated graphics may support understanding of route clearances.

The purpose of this exploratory study was to obtain pilot subjective feedback on the possible use of animation to present route and reroute information on the ND.

Background

Animation is a form of computerized graphics used to convey dynamic situation information. Humans have a tendency to perform inferences of causality, which implies that people may have an internal animation process in their mind (Hegarty, 1992). Researchers have shown that some animation applications have improved operator performance. Roscoe and Jensen (1981) showed that animated graphics of visual guidance and flight-prediction enhanced pilots' accuracy of curved approaches to airport runways. According to Payne, Chesworth and Hill (1992), a short animation to instruct how to use a computer program was effective. The properties of animation for performance improvement using animation have been investigated. Bartram (1997) theorized preattentative and interpretative properties of motion made animation a useful scheme to express complex information. Bederson and Boltman (1999) found that animation enhances operator's ability of reconstructing spatial information space without any degradation on task time. According to Tversky, Morrison and Betrancourt (2002), animated graphics can represent more information than static graphics or text, and animation enhanced the interactivity between operators and the display interface. An advantage of animation was shown for extracting incidental information in education applications (Rieber, 1991). In classes using computer-assisted education to teach complex and sophisticated concepts, animations were beneficial (Thatcher, 2006; Lin, 2011).

Researchers have also shown that animations were a perceptually problematic form, and their efficacy was questionable (Morrison, Tversky, & Betrancourt, 2000; Tversky et al., 2002). Bartram (1997) found that some operators were unable to correctly perceive the animation speed and this led them to dislike animations. Tversky et al. (2002) asserted that excessively complex and fast visual effects in animations used to convey the information of complex systems inhibited operators' ability to apprehend the information. If a given event is not inherently integrated into sequences of discrete steps in nature but a continuous event, the expressed animation may not deliver the right information (Morrison et al., 2000; Lee, Klippel, & Tappe, 2003). Researchers also found that animated displays are not effective because there were mismatches between the attributes of motion and the nature of the task at hand (Rieber, 1991; Morrison et al., 2000). Many animated displays often highlighted task-irrelevant features

(Lee et al., 2003). Morrison and Tversky (2001) showed both static graphics and animated graphics were preferred to text to convey spatial information, but the static graphics and animations did not show any significant difference in the spatial reasoning performance. Lee and Klippel (2005) found another pitfall of animated graphics for a spatial route finding application. Operators paid attention to the critical landmarks and non-critical landmarks equally while the animation was played (Lee & Klippel, 2005).

Researchers have investigated ways to make animations more effective. There are design guidelines for animation; "the animation process should be represented according to the user's mental model and system entities, and the user should be allowed to control the animation duration and replay it" (Stasko, 1993). Other examples of proposed principles for reducing the cognitive load caused by animations are as follows; animations should be segmented into small sections (Ayres & Paas, 2007), provide operators the capability of controlling the presentation (Ayres & Paas, 2007; Hegarty, 2011), signal the key information (Ayres & Paas, 2007), limit extraneous factors (Ayres & Paas, 2007), use of hybrid approaches of static and animated graphics in applications; animations accompanied with static diagrams are an effective alternative to animated-only instructional procedures in some applications (Lee et al., 2003; Ayres & Paas, 2007). Another proposed strategy for effective animations was adding attention cueing such as arrows, color, or luminance contrast (spotlight effect) to create instructional animations by integrating a series of static representations (de Koning, Tabbers, Rikers, and Paas, 2009; Amadieu, Marine, and Laimay, 2011). The fundamental idea of cueing is similar to the concept to designing salient features into static representations. Cues should be used to draw user attention to the key attributes illustrated in the animation within the animation time period. de Koning et al. (2009) provided a framework for cueing design; guiding operator attention to facilitate the selection and extraction of essential information, emphasizing the main idea of instruction and organization, and making clear the relations between and within elements to foster integration. These attention cues enhanced operator comprehension both of cued information and uncued information (de Koning, Tabbers, Rikers & Paas; 2007). de Koning et al. (2009) suggested developing new cues that work in animations rather than simply reusing the cues which had been effective for static representations for the same task.

The use of attention cueing described by Amadieu et al. (2011) and de Koning et al. (2009) were used to design animation into a ND with the idea of supporting pilot interpretation of DataComm clearances. This first study was developed to obtain pilot feedback related to the use of animation.

Evaluation

Prototype

A prototype and test system of an ND that presented both current route and cleared routes with animation was developed. Figure 1 presents a sample screenshot of the ND. The test system included a current flight plan overview section, a text (DataComm) clearance section, and a ND section. A magenta line was used to indicate the original flight path as it is currently used on NDs. A white triangle at the bottom of the display indicates ownship. Graphic symbols were also used to indicate ground stations such as VORs, airports, and waypoints. Novel graphics were created to present the cleared route (see Figure 2). Each cue was designed to guide the operators' attention to specific locations and detailed operation indications (de Koning et al., 2009; Amadieu et al., 2011). They included the following:

- Dotted green line: new flight path of ownship.
- Caret green line: new heading of ownship.
- Text label: new altitude of ownship.
- Ownship changes color based vertical clearance information, and moves along routes.

When activated, the ownship symbol moves along the new path indicated by the DataComm clearance. The test program was developed using JAVA language under NetBeans SDK version 7.0 and presented to pilots on a monitor. An external numeric keypad was used to start the animations. Half the route clearances were incorrect scenarios with respect to the flight plan and the other scenarios were correct.

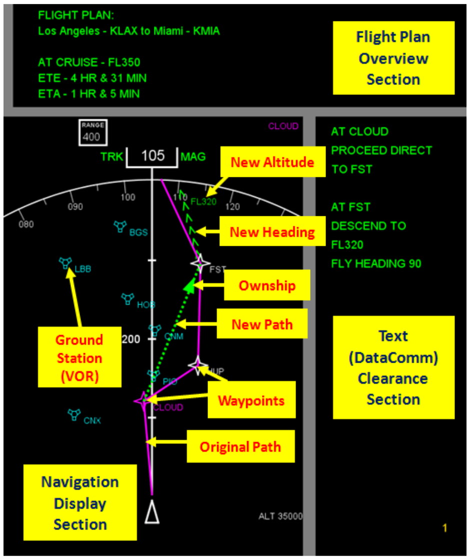


Figure 1. Setting of Test Program.

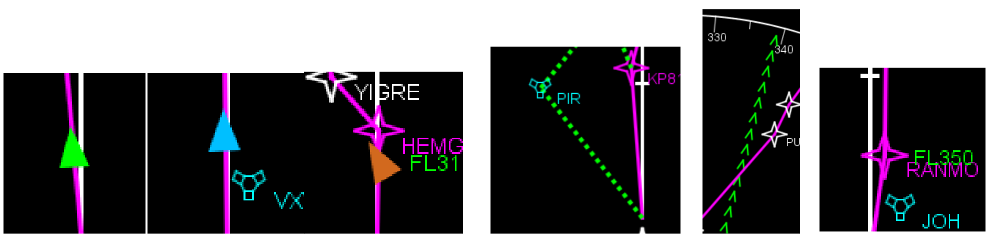


Figure 2. Ownship Symbol (Cruising, Climbing, and Descending), Dotted Line (New Flight Path), Caret Line (New Heading), Text Label (New Objective Altitude)

Rules of Test Animation

- The triangular ownship symbol moved over the new route. The new route could include part of the original route. All the composed clearances had sequential scenarios, but some clearances included simultaneous scenarios (e.g., "DESCEND TO FL320, FLY HEADING 90" as shown in Figure 1).
- Rather than drawing the entire route clearance as a static graphic first and having the ownship move over the graphics, the route was drawn as discrete steps similar to the way in which the clearance was worded as text. After a portion of the route was drawn, the ownship moved, followed by the drawing of the next part of the route clearance. This continued until the animation was complete. Each part of the reroute graphic remained on the screen after it was drawn. This provided the viewer with a “trail” of the ownship and allowed the pilot to visualize the entire route at the end of the animation.
- The ownship symbol was represented in three different colors depending on the ownship’s altitude status. When ownship was requested to climb, the color was sky blue. When ownship was requested to descend, the color was brown. When ownship was instructed to fly at cruising level at an altitude, the color was green. The color code of sky blue and brown was to suggest *sky* and *ground* similar to that used in a primary flight display (PFD). The ownship symbol paused for 0.5 seconds when it arrived at a waypoint or a specified altitude to represent discrete elemental changes as suggested by de Koning et al. (2009) to support relationships between elements and integration.
- The caret line is drawn when the ownship is directed to change heading. The caret lines were drawn from the ownship symbol to the spot on the circumference of the electronic compass in the ND for the angle indication. When the ATC request included a heading change, a caret line was drawn first and the ownship moved over the caret line to indicate that it moved over the path of new heading.
- The text labels of flight level as "FL350" indicated the location where the ownship should meet the altitude.

Figure 3 is an example of steps in an animation with an acceptable clearance scenario. Figure 4 is an example that would be obviously unacceptable.

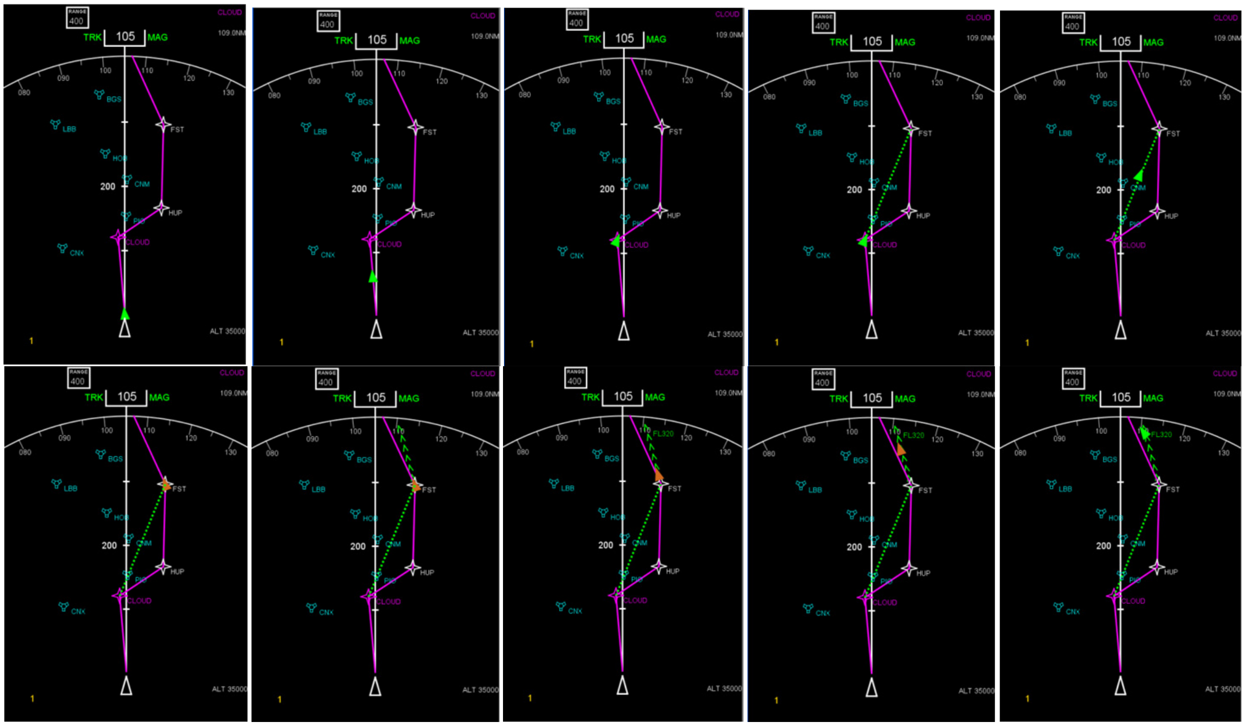


Figure 3. Steps of an Animation with Acceptable Clearance Scenario (Left to Right, and Up to Down)

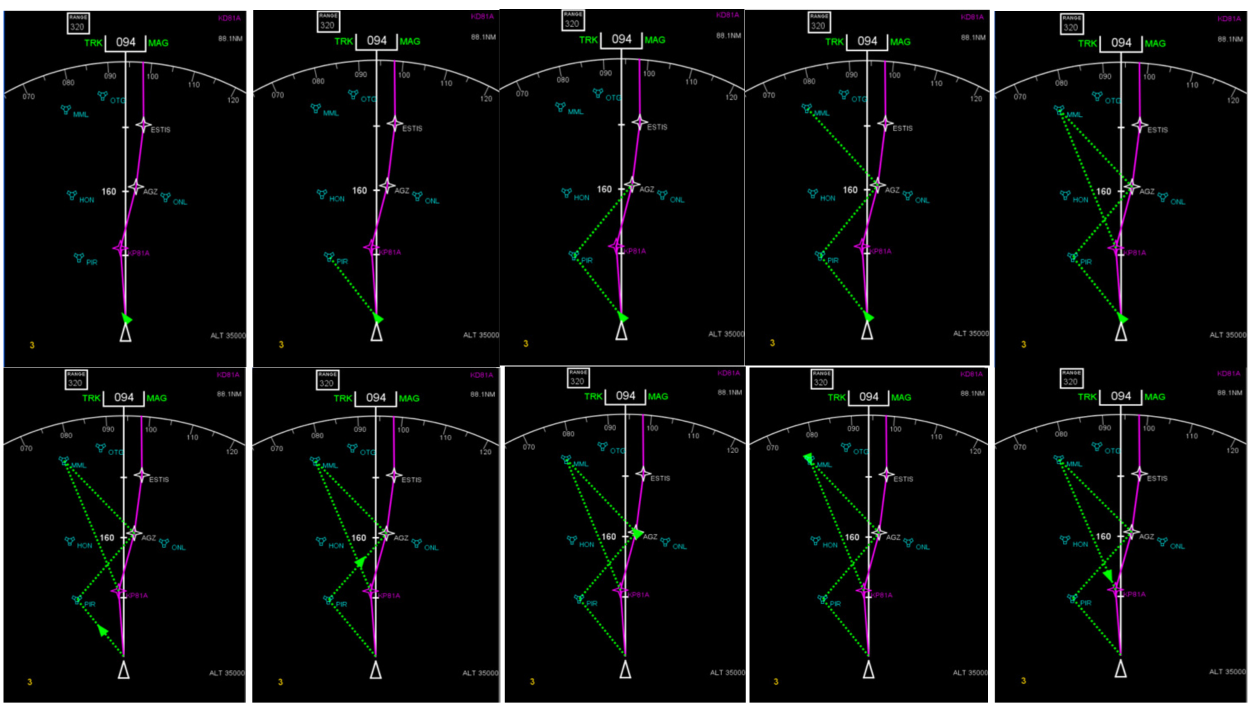


Figure 4. Steps of an Animation with an Unacceptable Clearance Scenario (Left to Right, and Up to Down)

Method

Twenty-five civilian pilots (all male) were asked to review the animation concepts. Subjects were briefed on the test system including the flight overview, the ND, the clearance information, and how to play an animation. They were asked to review a minimum of five scenarios. Subjects could ask questions and talk about the concepts as they progressed. Their comments were noted.

After reviewing a minimum of five scenarios pilots were asked their opinions about 1) whether animation would support their ability to correctly interpret the reroute clearances and 2) if animation would help them to rapidly understand the clearance compare to no animation. Subject answers were categorized into three levels 1) positive comment, 2) negative comment, 3) neutral.

Results

Most pilots were able to give their opinions by reviewing less than 10 scenarios. Figures 5 and 6 present the frequency counts for the three categories for both questions. Results showed that most subjects thought that animation would support the correct interpretation of clearances and support rapid understanding. Subject comments are summarized below.

- The animations would help them understand which portions of the clearance should be completed sequentially versus simultaneous.
- The animation created unnecessary effects that would distract them from desired actions.
- Add the ability to control the animation speed based on user preference. For complex clearances they may want to slow down the animation speed.
- Create a pause function so they can focus on specific regions.
- The subjects indicated that they might miss a color change if the ownship moved over a short segment so recommended changing the line colors to match the different ownship colors. They also suggested changing the color of the text clearance in the same way.
- The animations would be effective as an optional function, because of individual differences in understanding clearances.

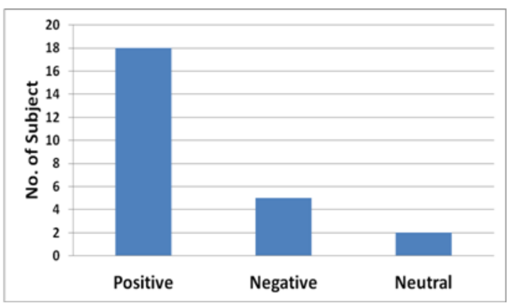


Figure 5. Subjects' Opinion of Animation to Support Correct Interpretation of Clearances

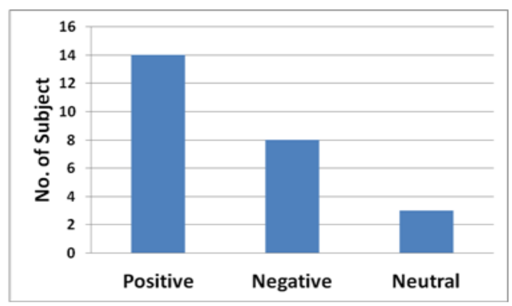


Figure 6. Subjects' Opinion of Animation to Support Rapid Understanding of Clearances

Discussion

The animated ND display to support understanding of complex text clearances was designed using the principles of attention cueing for animation proposed by de Koning et al (2009) and Amadieu et al. (2011). Most subjects thought that the animations would support rapid and correct understanding of text clearances. This was a first step toward moving to a more detailed design and testing. The drawbacks that pilots mentioned included the difference between their reasoning speed and the speed of the animation. For some, the animation may have been too slow and for others too fast. The speed may need to vary at different times within a complex clearance. If changes are happening simultaneously, the speed may need to be slower, while sequential changes may be presented at higher speed. Subjects' design recommendation of adding an animation speed control is consistent with the implications by Stasko (1993), Ayres and Paas (2007), and Hegarty (2011). To determine appropriate speed ranges additional research is needed. The pilots could then choose from a level of "best" ranges based on their preferences.

Using animation on the ND would require pilots to attend to the display for the period of time that the animation played. Given the need for pilots to multitask this may not be feasible for one-seat cockpits. However, the ability to replay an animation as often as needed may support the pilot so that they can review when they choose. They may also choose animation for specific situations, for example when the clearance is considerably complex or to support one area of the clearance that is not quite understood.

It is recommended that animation always be optional for pilots. However, there is a need for an in-depth study to determine recommended design parameters. The next step would be to evaluate pilot performance using objective measures of accuracy and time to respond to help select the design parameters. These data could be compared to the recent time and accuracy data for presenting text and graphic DataComm clearances by Gallimore et al. (2013).

References

Amadieu, F., Mariné, C., & Laimay, C. (2011). The attention-guiding effect and cognitive load in the comprehension of animations. *Computers in Human Behavior, 27*(1), 36-40.

Ayres, P., & Paas, F. (2007). Can the cognitive load approach make instructional animations more effective? *Applied Cognitive Psychology, 21*(6), 811-820.

Bartram, L. (1997). Perceptual and interpretative properties of motion for information visualization. *Proceedings of the 1997 Workshop on New Paradigms in Information Visualization and Manipulation,* 3-7.

Bederson, B. B., & Boltman, A. (1999). Does animation help users build mental maps of spatial information? *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on,* 28-35.

de Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2007). Attention cueing as a means to enhance learning from an animation. *Applied Cognitive Psychology, 21*(6), 731-746.

de Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review, 21*(2), 113-140.

Gallimore, J.J. Kiss, S.B., Munoz, R.D., Oh, C., Crory, T., Ward, B., Green, R., Shingledecker, C., and Tsang, P. (2013). DataComm – Display Alternatives for the Flight Deck: Overview and Human Factors Recommendations. Final Tech Report Vol 1 & 2. FAA NextGen Advanced Concepts and Technology Development, Human Factors Division (ANG-C1), DTFAWA-10-A-80021. Wright State University, Dayton OH, 45435. Retrieved from http://www.hf.faa.gov.

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 1084.

Hegarty, M. (2011). The cognitive science of Visual-Spatial displays: Implications for design. *Topics in Cognitive Science, 3*(3), 446-474.

Lee, P. U., & Klippel, A. (2005). Dynamic aspects of spatial information in air traffic controller displays. In T. Barkowsky, C. Freksa, M. Hegarty & R. Lowe (Eds.), *Reasoning with mental and external diagrams: computational modeling and spatial assistance* (p. 18–23). Stanford: AAAI Press.

Lee, P., Klippel, A., & Tappe, H. (2003). The effect of motion in graphical user interfaces. *Smart Graphics,* 277-290.

Lin, H. (2011). Facilitating learning from animated instruction: Effectiveness of questions and feedback as attention directing strategies. *Journal of Educational Technology & Society, 14*(2), 31.

Morrison, J. B., & Tversky, B. (2001). The (in) effectiveness of animation in instruction. *CHI'01 Extended Abstracts on Human Factors in Computing Systems,* 377-378.

Morrison, J. B., Tversky, B., & Betrancourt, M. (2000). Animation: Does it facilitate learning. *AAAI Spring Symposium on Smart Graphics,* 53-59.

Payne, S. J., Chesworth, L., & Hill, E. (1992). Animated demonstrations for exploratory learners. *Interacting with Computers, 4*(1), 3-22.

Rieber, L. P. (1991). Animation, incidental learning, and continuing motivation. *Journal of Educational Psychology, 83*(3), 318.

Roscoe, S. N., & Jensen, R. S. (1981). Computer-animated predictive displays for microwave landing approaches. *Systems, Man and Cybernetics, IEEE Transactions on, 11*(11), 760-765.

Stasko, J. T. (1993). Animation in user interfaces: Principles and techniques. *User Interface Software,* 81-101.

Thatcher, J. D. (2006). Computer animation and improved student comprehension of basic science concepts. *JAOA: Journal of the American Osteopathic Association, 106*(1), 9-14.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*(4), 247-262.

MANNED-UNMANNED TEAMING: TRAINING US ARMY UNMANNED AIRCRAFT SYSTEM OPERATORS IN THE SCOUT-RECONNAISSANCE ROLE

John E. Stewart
US Army Research Institute
Fort Rucker, AL

Until recently, U.S. Army unmanned aircraft systems (UAS) were intelligence-gathering platforms. The UAS mission has recently changed from strategic intelligence, surveillance, and reconnaissance (ISR) to scout-reconnaissance (SR) operations. This shift has produced an increased requirement for coordination between manned and unmanned aircraft. Manned-unmanned teaming (MUM-T) requires that UAS operators become knowledgeable and proficient in the same scout-reconnaissance (SR) skills as pilots of armed helicopters. This paper summarizes the many training challenges consequent to the move from ISR to SR roles. It will review completed and ongoing research efforts by the Army Research Institute (ARI) Fort Rucker element, which investigated (a) preparedness of UAS operators to perform tactical SR missions, (b) the training provided at a Combat Training Center (c) differential perspectives of the manned and unmanned Army aviation communities on the role of UAS in MUM-T, (d) MUM-T skills that must be trained and measured.

Background

Shadow (RQ-7B) is a medium unmanned aircraft system (UAS) and currently the most numerous in the US Army inventory. Other medium to heavy Army UAS are also employed as scout-reconnaissance (SR) assets. These are the MQ-1C Gray Eagle and the MQ-5B Hunter. The Army UAS mission until recently was intelligence, surveillance and reconnaissance (ISR) in which UAS operators proceeded to a predetermined location, collected data, and stayed within the assigned grid until instructed to proceed to another location. Real-time video feed was provided to the Brigade Tactical Operations Center where image analysis was performed. The new SR role is quite different. The tasks include zone and route reconnaissance, in which the RQ-7B flies in increasing, concentric circles to assure that the area is free of potential threats, as well as laser designation and handover of targets to armed helicopters. Although route and zone reconnaissance superficially resemble ISR tasks, they are very different. UAS operators must understand, interpret and develop the tactical situation. *Developing the situation* entails communicating and coordinating with ground forces and/or other aircraft, attribution of target intent, and determining impact of the situation on friendly forces. These SR skills are traditionally performed by crews of scout helicopters, such as the OH-58D.

Manned-Unmanned Teaming (MUM-T)

One advantage of training UAS operators in the SR role is better coverage of the battlespace by combining the complementary strengths of armed helicopters and UAS. The OH-58D and AH-64D can find, designate and destroy targets under day or night conditions, but have limited endurance and typically operate at low altitudes. The RQ-7B, also equipped with a sensor suite (mission payload) can remain airborne for over 6 hours and operate above 6,000 feet, giving it a different vantage point to detect and identify targets, and report changes in the tactical situation to the armed helicopter. The helicopter can attack the target from a more covert location, sometimes without seeing it. MUM-T missions, a result of the shift from ISR to SR, will impact training requirements for the UAS air vehicle and mission payload operators. This paper reviews recent Army Research Institute (ARI) efforts to understand the training implications of these changes.

Current Status of UAS Operator Training

Proficiency of New UAS Operators Reporting to the Unit

ARI asked the principal staff officers (Majors and Lieutenant Colonels) at three selected Brigade Combat Teams (BCT), the extent to which new UAS operators reporting to the BCT were able to perform critical SR tasks (Stewart, Bink, Barker, Tremlett, & Price, 2011). The staff officers interviewed comprised Intelligence (S2), Operations (S3), and Brigade Aviation officers (BAO), as well as senior leaders and trainers of two UAS Platoons (each BCT has one MI Company, which includes a UAS Platoon). Of the three staff officers, only the BAO is an aviator. The S2 and S3 are ground officers, who plan the RQ-7B missions. Interviews revealed that the perspective of the S2 tended to be military intelligence (MI) in terms of UAS employment, in which the mission was ISR. Neither the S2 nor the S3 showed extensive knowledge of how to employ UAS as an SR asset, though they agreed that new UAS operators required additional training at the BCT before they could execute SR missions. BAOs and senior UAS Platoon members had greater knowledge of aviation tactical operations and generally believed that UAS operators lacked the necessary SR skills when reporting to the BCT.

Thus the consensus of respondents was that RQ-7B operators arrive at the unit poorly trained for tactical SR mission execution, except for Airspace Operations, which are well-trained institutionally at Fort Huachuca, AZ. The ARI research team also learned that opportunities for unit level training and practice of SR skills while the BCT is at home station are limited due to multiple factors, including airspace restrictions on pilotless aircraft, cost, equipment availability, and inadequate range areas. Respondents added that Combat Training Centers (CTC) provided the only chance for unit-level collective training and practice while still in the United States. Because of limited training opportunities at home station, the most effective SR training must take place on the job while the BCT is deployed.

UAS Operator Collective (Unit Level) Training at a Combat Training Center

The Joint Readiness Training Center (JRTC), Ft Polk, LA, is a CTC where operational units perform live training exercises, facing a skilled role-playing opposing force. JRTC has dedicated senior leaders and Trainer-Mentors (TM) who provide training and feedback to unit leaders and personnel. The ARI research team interviewed TMs at Brigade, Battalion, and Company levels (Stewart, Barker, & Bink, 2010). One most important finding was that UAS operators at JRTC were not evaluated on criteria relevant to performance of the SR mission. Instead, UAS teams were evaluated on launch, operation, and recovery of the aircraft, as well as total hours flown, and often had no clear indication of task and purpose of the mission, (e.g., check on the status of the troops). TMs noted that staff officers at Battalion and Brigade level did not seem to have a strong background in UAS operations. There were indications of the survival of MI culture: the UAS mission plan was still called the (data) collection plan. The S2, sometimes assisted by the S3, usually took the lead in planning the mission. Trainers and leaders at JRTC remarked that the leadership of the BCTs often required coaching in how to employ the RQ-7B as a mission asset, and recommended that education in UAS utilization be elevated all the way to command level.

Roles and Critical Skills Required of UAS Aircrews for MUM-T

Perceived Roles and Status of UAS and Manned Aviation in MUM-T

Stewart, Roberts, and Bink (2012) conducted a survey of 34 US Army helicopter pilots and 31 Army UAS operators. The questionnaire consisted of two parts: (a) present and future effectiveness of UAS in the SR role, (b) which of eight selected SR mission tasks were most appropriately performed by manned aircraft, UAS alone, or UAS and manned aircraft (MUM-T). UAS and manned respondents agreed that the role of UAS in the SR operations would expand. UAS personnel indicated that UAS would eventually perform many if not all of the SR mission tasks currently performed by armed scout helicopters. By contrast, helicopter pilots indicated that the role of UAS would be to assist, not supplant, armed helicopters in the SR role, (see Table 1).

Table 1.

Perceptions of Future Roles of Manned and Unmanned Aircraft

Item	Subgroup	%Agree	% Disagree
UAS will assume a more active role in SR mission.	Manned	97	3
	Unmanned	90	10
UAS operators will have to learn to develop the situation once the target is identified.	Manned	94	6
	Unmanned	83	17
UAS will eventually become an equal status player in SR operations.	Manned	29	71
	Unmanned	77	23
UAS has made significant contributions to manned helicopter SR operations.	Manned	88	12
	Unmanned	74	26
In the future, UAS will completely take over the SR role in tactical operations	Manned	6	94
	Unmanned	52	48
UAS operators must assume a more active role in SR than merely providing an airborne sensor.	Manned	77	23
	Unmanned	97	3
Replacement of the OH-58D and AH-64D by weaponized UAS is unrealistic notion.	Manned	76	24
	Unmanned	27	73

UAS operators, presented with a list of eight SR mission tasks, indicated that UAS could perform most of these tasks. Pilots saw most tasks as appropriate to manned aircraft. Even with these differences in perceptions of UAS roles, it still appeared that most of the eight mission tasks presented were appropriate to combined manned and unmanned operations. Thus the majority of respondents saw each of the eight SR mission tasks as appropriate to both manned aircraft and UAS, though patterns of responses tended to differ for these two groups (Table 2).

Table 2.

Perceived Appropriateness of Selected Mission Tasks for Manned-Unmanned Team Operations

Mission tasks UAS Respondents considered appropriate for both manned aircraft and UAS.	Actions on Contact Downed Aircraft Recovery Fundamentals of Security
Mission tasks Manned Respondents considered appropriate for both manned aircraft and UAS.	Aerial Observation Fundamentals of Reconnaissance
Mission tasks both UAS and Manned Respondents considered appropriate for both manned aircraft and UAS.	Laser Target Handoff to Ground Target Handover SALT-W *Reports

Note: Size, Activity, Location, Time, What (procedure for reporting targets/activities observed).

The final pattern of responses showed both manned and unmanned respondents likely to indicate three mission tasks as appropriate for both UAS and manned aircraft. The tasks that represent this pattern of responses were Laser Target Handoff to Ground, SALT-W Reports, and Target Handover. Of all eight tasks, the outlier seems to be Actions on Contact, which 62% of manned respondents perceived as appropriate primarily to manned aircraft. Similarly, though 88% of manned respondents saw Aerial Observation as a mission task for both UAS and manned aircraft, unmanned respondents were split evenly with 48% stating it was primarily a UAS mission, and 48%, a mission for both aircraft types.

These findings provide important feedback to decision makers regarding the perceived present and future tactical roles of manned and unmanned aircraft. Knowing the current attitudes toward capabilities of UAS could provide insight for training developers who must devise strategies for training manned and unmanned aircrews to work together as players in MUM-T. The findings also point to the need to specify more precisely the respective roles of manned and unmanned team members before UAS can fully participate in MUM-T.

Identifying and prioritizing critical MUM-T skills

Sticha, Howse, Stewart, Conzelman, and Thibodeaux (2012), used a method similar to the Air Force Mission Essential Competencies (MEC) approach (Colegrove & Bennett, 2006) to identify and prioritize the most critical individual skills supporting MUM-T, and to pinpoint performance indicators for these skills. The investigators began with a review of Army doctrinal materials to identify (a) missions in which UAS operators would need to coordinate with helicopter pilots, (b) specific tasks required to perform these missions, and (c) MUM-T skills involved in executing these tasks. Training-critical skills were defined by two criteria: (a) inadequate performance would lead to mission failure or serious risk to personnel and/or equipment, (b) UAS operators recently graduated from training do not possess these skills.

Three workshops were conducted with small groups of doctrine developers, scout-attack helicopter pilots, and UAS senior instructor-operators to determine the relative importance of the skills as to training criticality, and to derive behavioral indicators of performance. The list of 25 skills was confirmed as relevant to MUM-T and doctrinally correct in a focus group attended by UAS operators, scout-attack helicopter pilots, and doctrine developers. Ratings and rank-orderings by participants indicated that all 25 skills were at least moderately important to SR missions, and present serious risks if performed incorrectly. Perceived levels of competency of UAS operators to perform the skills varied greatly, indicating that many were not addressed in training. Some skills, though highly important, were not rated as training-critical because they were adequately trained. Table 3 presents the 10 most training-critical skills.

Table 3.

Rank Ordering of Ten Most Training-Critical MUM-T Skills

Rank	Skill
1	Deconflict munition trajectories from airframe.
2	Utilize standard execution commands to initiate attack.
3	Transmit information on method of attack.
4	Switch roles of laser designator.
5	Conduct call for direct fires.
6	Select best weapon system.
7	Develop/send common operating picture information.
8	Utilize joint, Army, & civilian personnel recovery terminology.
9	Prioritize engagement of targets.
10	Gain and maintain enemy contact.

The final list of MUM-T skills was reduced to 20 (some were redundant with others, or judged as holdovers from ISR). For these, 140 performance indicators were derived. These skills and indicators support the development of two ongoing ARI efforts: a PC-based training tool for practicing the identified cognitive and procedural skills in the unit, and benchmarked performance measures, designed for use in networked, virtual environments, as well as live field exercises, where much of MUM-T training will take place. Table 4 presents an example of behavioral indicators for the skill ranked most important of all 20 remaining skills.

Table 4.

Performance Indicators for Skill: Deconflict Munition Trajectories from Airframe

Call for fire is complete and accurate.
Operator is aware that deconfliction of airspace is taking place.
Operator is aware of positions of friendly assets (e.g., aircraft and ground units).
Operator confirms when (friendly aircraft and/or ground assets) clear.
Operator determines if rounds are accurately placed on target.

Discussion

It is evident that much needs to be done in the development of SR training for the Army UAS operator. Most tasks that operators will be called on to perform in the operational unit are not those acquired institutionally during advanced individual training. These are tactical skills, supporting unit-level mission tasks that typify Army scout-attack and reconnaissance operations. Most SR training for UAS operators must take place in the operational unit at home station, or during deployment. Most collective MUM-T training at home station will of necessity have to take place in networked, shared virtual environments. In addition to knowing what skills are most critical, it will also be necessary to develop behaviorally-anchored performance measures for each skill. This will assist trainers and Company commanders in assessing efficiency and effectiveness of collective MUM-T training. Individual training of cognitive and procedural skills not requiring simulators can be executed on new-generation portable training devices, such as Tablet PCs. The technology of handheld devices is evolving rapidly; hence, it is likely that training apps allowing for wireless networking will allow some degree of team-level practice on these devices at home station.

These technologies for implementation exist at present, but the real challenge to implementation may be due to differences in organizational culture. Army Pilots and UAS operators come from quite different backgrounds in terms of formal education, training, required aptitudes and rank. UAS operators are enlisted personnel and noncommissioned officers. Pilots are either commissioned or warrant officers. Our research efforts have shown that perceptions by members of these two groups of the relative roles and capabilities of UAS and manned aviation differ in many ways. Senior UAS instructor-operators generally believed that UAS air vehicle and mission payload operators could acquire the critical skills necessary to execute most if not all of the SR tasks that armed helicopter crews can perform. By contrast, most manned helicopter pilots perceived UAS as helping the "shooter" find and attack targets. This is an important role, but nonetheless subordinate to manned aviation. Among the open-ended statements by senior members of the UAS community was that UAS can successfully assume the tactical role, if the manned community would allow it to do so. Looking at the present state of UAS operator training, it is apparent that newly-trained UAS operators cannot perform the most critical of the SR skills called for by the MUM-T mission, for the simple reason that these are not trained as part of their common core and aircraft qualification training. The explanation of this is two-fold: (a) institutional (i.e., schoolhouse) training time is limited, and there is little opportunity to learn SR skills during primary and advanced individual training, and (b) a vestige of MI climate is still evident at the institutional phase of training. Stewart, et al. (2011) found that unit leaders and trainers believed that more training in SR fundamentals could take place at the institution. In short, more institutional training in the foundations of SR would enhance preparation of UAS operators for training in the unit, but more opportunities for home station training would also have to be provided at the unit. The training institution at Fort Huachuca is striving to revise its curriculum to include more material pertinent to SR operations, using multimedia approaches tailored to the cognitive processes of student operators. Stewart, Roberts, and Bink (2012) have suggested that UAS aircrews spend part of their training time at the Army Aviation Center at Fort Rucker, AL, planning and rehearsing simulated missions alongside instructors and students from the scout-attack helicopter community. This would also serve another purpose of integrating UAS into aviation training. The formation of UAS-manned aviation teams is being facilitated by the stand-up of the Full Spectrum Combat Aviation Brigade (FSCAB), starting with the 101st at Fort Campbell, KY. The 101st has one Battalion of OH-58D and RQ-7B aircraft, whose mission is to execute cooperative engagements as manned-unmanned teams. By requiring UAS and manned helicopter crews to interact during the mission planning and execution processes, this will likely be the initial step in successfully assimilating UAS into the world of tactical Army Aviation.

Acknowledgments

We are grateful to the Training and Doctrine Command Capabilities Managers for UAS and Reconnaissance-Attack, as well as the Directorate of Training and Doctrine at Fort Rucker, AL, for their support and assistance at all phases of this research, especially for the superb quality of subject-matter expertise provided. We thank the intelligence, operators, and Brigade aviation officers from three unnamed Brigade Combat Teams, and members of two RQ-7B Platoons for their valuable insights on unit level training of UAS aircrews. We are indebted to military and civilian personnel from various Army installations in the United States for their participation in the UAS-manned aviation survey.

References

Colegrove, C. M., & Bennett, W. (2006). *Competency-based training: adapting to warfighter needs*. (AFRL -HE-AZ-TR-2006-14). Air Force Research Laboratory: 711th Human Performance Wing.

Stewart, J. E., Barker, W. C., & Bink, M. L. (2010). *Army RQ-&B unit training issues*. Paper presented at 63rd meeting of Department of Defense Human Factors Engineering Technical Advisory Group, Tempe, AZ.

Stewart, J. E., Bink, M. L., Barker, W. C., Tremlett, M. L., & Price, D. (2011). *Training needs for RQ-7B unmanned aircraft system operators in the scout-reconnaissance role* (Research Report 1940). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Stewart, J.E., Roberts, K.R., & Bink, M.L. (2012). *Unmanned aircraft systems in the scout-reconnaissance role: perceptions of the U.S. Army manned and unmanned aircraft communities* (Research Report 1956). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Sitcha, P.J., Howse, W.R., Stewart, J.E., Conzelman, C.E., & Thibodeaux, C. (2012). *Identifying critical manned-unmanned teaming skills for unmanned aircraft system operators*. (Research Report 1962). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

MILITARY UNMANNED AIRCRAFT SYSTEM OPERATORS: TRAINING AND HUMAN PERFORMANCE ISSUES

CDR Joseph V. Cohn, MSC, USN, Ph.D.
Office of Naval Research
Arlington, Virginia
LCDR Brent A. Olde, MSC, USN, Ph.D.
Naval Air Station Patuxent River
Patuxent River, Maryland
Dr. Richard D. Arnold
Naval Medical Research Unit - Dayton
Ms. Elizabeth B. O'Neill
Strategic Analysis, Inc.
Arlington, Virginia

Unmanned Aviation System (UAS) have leveraged considerably from the manned aviation approach. This approach was useful to jumpstart this technology but it is now time to find a more efficient and up-to-date approach, keeping with the capabilities and limitation of the concurrent technology. Three technical areas to mature and address the current limitations are recommended: 1) Selection for UAS Personnel (SUPer), to accurately forecast candidate UAS operator performance across UAS platforms and missions; 2) Distributed, Adaptive & Modular entities for UAS (DyAdeM), to automatically generate realistic & adaptive synthetic environments for simulated UAS training; and, 3) UAS Control Station Human Machine Interface (CaSHMI), to provide validated information display concepts.

Beginning in FY12, the US Navy and Marine Corps began to significantly increase their rate of acquisition of a wide range of Unmanned Aerial System (UAS) platforms. The Army and Air Force have already fielded thousands of these systems in the past decade and plan to continue to do so. These UASs will include significant technological advances, blending automation with dynamic, decentralized control, and will support a wide range of missions. Yet, despite these technological advances and increases in automation, these systems do not lend themselves to easy use by their operators. From a human systems integration perspective, they are not well-supported by control station interfaces, training technologies, or selection tools; as indicated by the fact that as much as 50 percent of all UAS mishaps are attributed to human factors (Williams, 2004; Tvaryanas, 2005, 2006).

Reducing mishaps and unsuccessful UAS operations will require better interface design and a new kind of operator - one who has been specifically selected, trained and equipped to process information safely and effectively. They must interact with cutting-edge technologies, work collaboratively with others, and effectively manage their cognitive workload and attention over long mission durations (McCarely & Wickens, 2005). Since the 1990s, all three services have conducted varying levels of investigation to better understand how to select Air Vehicle Operators (AVOs), how to train them, and how to equip them. These early studies often assumed that selection and training criteria should be similar to those used in manned aviation (Barnes, et al, 2000; Hall & Tirre, 1998) and that manned aviation was the gold standard against which UAS AVO requirements should be considered. Many of these early studies focused on assessing UAS operator requirements for platforms whose control schema and missions mirrored those of manned aviation platforms (Biggerstaff et al, 1998; Kay et al, 1999). The UAS landscape has changed with fiscal realities and evolving mission sets, combined with significant advances in the state of the art of UAS platforms and control schema, which have all evolved the AVO into a mission manager vice a hands-on controller. Today, simply replicating the manned aviation select-train-equip approach is an inefficient solution at best and a potential disaster at worst (McCarley & Wickens, 2005). As stated by a recent US Air Force Scientific Advisory Board: "the considerable base of human factors knowledge derived from cockpit experience may have limited applicability to future systems…" (US Air Force, 2004).

While this sentiment speaks to future aviation systems at large, UASs are a significant departure from traditional roles and responsibilities for its human operator (Figure 1). First, the AVOs, unlike manned aviators, are not co-located with their platform. This decoupling of the human from the system has created unique human system integration issues (Tvaryanas, 2006). Compared to their manned aviation counterparts, AVOs work in sensory-deprived conditions, lacking the visual, auditory, and tactile cues present in manned aviation. Second, as automation becomes more reliable, the role of the AVO will continue to shift towards mission management, likely of multiple and different UAS platforms (Tvaryanas, 2006). Even today, AVOs, especially for larger more capable UASs, interact with their systems more through decision making, course of action planning, collaborative planning, and resource management than through hands-on 'stick and rudder' skills (Kay, et al., 1999). These roles and responsibilities are more reflective of mission management activities, like those of an Air Traffic Controller.

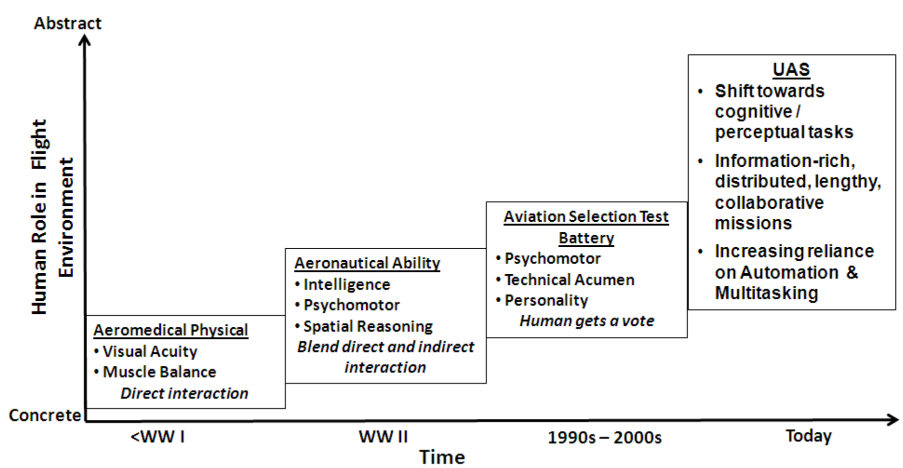


Figure 1: The Evolving Role of Human in Aviation Systems. In the early days of aviation, aviators had a very direct connection with their platform – basically a one-to-one mapping between the control inputs they provided and the subsequent behaviors the aircraft responded with. As aircraft became more complex, this mapping changed. With these changes came a shift in the role of the aviator, with a corresponding shift in the KSAOs they need in order to succeed. Today, UAS represent the most radical shift yet in this role, with the operators being 'dis' located from their platform; their inputs have very limited direct impact on their

aircraft's moment-by-moment action and mission success depends more on individual operators' ability to process and act on information and coordinate with other team members.

These, and other, differences between manned and unmanned aerial systems require new technologies and methodologies to ensure effective UAS operations. One approach is to address the selection, training, and interface design gaps as equal parts of a common approach for optimizing Human System Integration. This will require a detailed examination of: the types of Knowledge Skills and Abilities (KSAs) necessary to succeed in UAS operations; the best approaches for utilizing simulation-based training; and design approaches that will provide UAS operators with an effective way of interacting with their systems. The remainder of this paper discusses how the different Services may address these three areas.

Selection

Effective selection procedures identify individuals who possess a minimum level of qualifications and aptitude in the relevant knowledge, skills, and abilities to perform specific tasks and missions. Done properly, selection and classification procedures match overall training and interface design requirements. For manned aviation, the Department of the Navy uses a secure, web-based test delivery platform called the Automated Pilot Exam (APEX) to deliver the manned Aviation Selection Test Battery (ASTB) worldwide. APEX is government-owned and is capable of delivering psychomotor evaluations, tests of divided attention, and reaction-time evaluations using stick-and-throttle inputs or keyboard and mouse. It is also capable of administering computer-adaptive multiple-choice tests, which tailor test content to the examinee's ability level; this reduces test length, increases score accuracy, and greatly improves test security.

UAS platforms represent a unique domain from the perspective of the KSAs needed for successful operations. This is partly due to the wide range of platforms and missions that UAS support and partly due to the unique role that UAS operators are asked to assume. UAS operators must excel at integrating information from partial, incomplete, and abstracted data, attained from multiple sources, in collaboration with other UAS operators who may be located across vast geographical and temporal ranges. In the near future, they may also be required to concurrently operate more than one platform and perform more than one mission goal. As the technical capabilities of these platforms continue to grow, along with the mission sets, these cognitive and social competencies will have a far greater influence on mission success than the traditional ones currently selected for in manned aviation.

There are currently no tools in place to select and classify candidate UAS operators based on these competencies. Preliminary research suggests that such tools should include an emphasis on assessing: spatial capabilities (McKinley et al, 2011); social and interpersonal abilities and personality traits (Kay et al 1999; Carretta & Ree, 2003); executive processes, like attention management, information processing, multitasking, and decision making (Squire & Parasuraman, 2010; McKinley et al 2011); and human-autonomy interactions (McCarley & Wickens, 2005; Squire & Parasuraman, 2010). Similarly, there are no standards, policies, and guidelines for developing UAS operator career fields based on these competencies. Manpower and personnel decisions regarding the candidate pool from which AVOs are chosen have historically been based on manned aviation requirements rather than on UAS operational requirements. Recent studies conducted by the Air Force suggest that non-aviators may be as competent as aviators in terms of many of the KSAs associated with UAS operations (McKinley, et al., 2011).

Current efforts focus on developing a series of assessments to identify individuals with appropriate UAS AVO-relevant aptitudes and KSAs from populations including both military (Enlisted and Officer) and civilian personnel, as well as individuals with or without previous manned flight experience. The expected results from these efforts include the identification of KSAs and behaviorally anchored proficiency requirements for UAS AVO operators; the degree to which each KSA can be satisfied using civilian, Enlisted, or Officer candidates, and whether those candidates should have prior manned flight experience; and an identification of KSAs best attained through a training curriculum, along with guidelines for how to structure such a curriculum.

Training

Typical simulation-based training for aviation requires the integration of hundreds, if not thousands, of simulated entities into the overall training scenario. Developing these entities requires significant time and effort and results in entities whose behaviors are strictly guided, scripted, and limited based on pre-determined rules that define the entities' behaviors over the course of the training scenario. The net result is entities whose behaviors are not realistic, leading to reduced training effectiveness; yet this training requires significant effort to create, thus having prohibitively high authoring costs.

An alternative approach is to replace hand-coded rule sets with a capability to automatically generate new and appropriate Computer Generated Forces (CGF) behaviors from one or more data sources including: data captured during live UAS exercises or data captured as experts operate their systems within a simulated environment. On the basis of one or more of these initial data sets, it should then be possible to model those behaviors and provide new behaviors that will drive CGF entities in a training environment. This approach will require integrating cognitive modeling approaches with machine learning techniques to generate tactically authentic behaviors. Recent advances in the development of knowledge structures (Bermejo, 2006; Koeing, 2009) provide a formal approach for representing and characterizing underlying behaviors from large data sets (Boyce & Pahl, 2007), making it possible to capture structured data from multiple sources. Cognitive models provide a means of formally representing these underlying behaviors of interest. Machine learning techniques provide a wide range of inductive approaches to generalize these behaviors to new missions and contexts. Training objectives, doctrine and tactics, techniques and procedures (TTPs) bound the initial cognitive models and subsequent machine learning generalization to ensure that new behaviors are tactically authentic. The resultant behaviors can then be integrated into new training scenarios.

Current efforts focus on providing the underlying behaviors that drive CGFs. Of particular interest are behaviors driving large numbers of entities that provide the ecological background against which the “Patterns of Life” play out in the scenario. These include seemingly random actions of groups of individuals, ground vehicles, or surface ships as they affect and are affected by the trainee's actions. The manner in which these behaviors drive CGF, as well as the manner in which these forces are represented to the trainee, should be part of the design developed.

Interface Design

Critical challenges with using a single system to display information relating to operating multiple and different types of UAS platforms include: characterizing the necessary information that operators must interpret to make effective decisions; providing information in a way that allows for task switching and multi-tasking without reducing operator performance; enabling AVOs to manage the flow of information from UASs with varying levels of autonomy; supporting collaboration with other UAS teams and support personnel; and designing flexibility into the system to account for new platforms, missions and advances in information display technologies - such as those that adapt information based on context, mission, and user performance. At the core of this challenge lies the need to find platform-common information and platform-specific information requirements. These requirements should consider the mission characteristics required to perform in a “Patterns of Life” scenario. Once identified, follow the manned aviation lead in developing a “common” approach to representing basic aviation information (Wiener & Nagel, 1988; Mejdal, McCauley & Beringer, 2001). Approaches for representing information which will optimize AVO performance should be developed and design guidelines and solutions should be implemented similarly to other mission management-like domains, such as Air Traffic Control (Friedman-Berg, Yuditsky, & Smith, 2004). Traditional human factors techniques, (e.g., Wickens & Hollands, 2000) as well as more recently developed neuroergonomic assessment methodologies (e.g., Parasuraman & Rizzo, 2007), are expected to form the basis for much of this technical area.

Current efforts focus on documenting human factors-driven design guidance developed, in coordination with the appropriate Navy leads, for the Common Control System; CGF improvements to AVO training; and inputs and recommendations for KSAs to be part of AVO candidate selection and classification. For test and evaluation purposes, the Navy will make use of simulated Common Control stations, populated with CGFs and appropriate mission scenarios.

References

Barnes, M.J., Knapp B.G., & Tillman B.W. (2000). Crew systems analysis of unmanned aerial vehicle (UAV) future job and tasking environments. Aberdeen Proving Ground, MD: Army Research Laboratory, Report No.: ARL-TR-2081.

Biggerstaff S., Blower D.J., & Portman C.A. (1998). The development and initial validation of the unmanned aerial vehicle (UAV) external pilot selection system. Pensacola, FL: Naval Aerospace Medical Research Laboratory; Report No.: NAMRL-1398

Boyce, S., & Pahl, C. (2007). Developing domain ontologies for course content. Educational Technology & Society, 10 (3), 275- 288.

Carretta, T. R., & Ree, M. J. (2003). Pilot selection methods. In B. H. Kantowitz & P. S. Tsang & M A Vidulich (Eds.). Human factors in transportation: Principles and practices of aviation psychology (pp. 357-396). Mahwah, NJ: Erlbaum.

Friedman-Berg, F., Yuditsky, T., & Smith, A. (2004). Developing human factors design principles for information display systems in air traffic control. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48, 2386-2390.

Hall E.M., & Tirre W.C. (1998). USAF air vehicle operator training requirements study. Mesa, AZ: Air Force Research Laboratory; Report No.: AFRL-HE-BR-SR-1998-0001.

Kay, G., Dolgin, D., Wasel, B., Langelier, M. & Hoffman, C. (1999). Identification of the Cognitive, Psychomotor, and Psychosocial Skill Demands of Uninhabited Combat Aerial Vehicle (UCAV) Operators. Naval Air Warfare Center Report, retrieved 12 Feb 2012 from http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA368578.

McCarley J.S., & Wickens C.D. (2005) Human factors implications of UAVs in the national airspace. Atlantic City, NJ: Federal Aviation Administration, Department of Transportation; Report No.: AHFD-05-05/FAA-05-01.

McKinley, R.A., McIntire, L.K., & Funke, M.A. (2011). Operator selection for unmanned aerial systems: comparing video game players and pilots. Aviat Space Environ Med 82:635 - 42.

Parasuraman, R., & Rizzo, M. (2007). Neuroergonomics: The Brain at Work. New York: Oxford University Press.

Squire, P. N. & Parasuraman, R. (2010) 'Effects of automation and task load on task switching during human supervision of multiple semi-autonomous robots in a dynamic environment', Ergonomics, 53: 8, 951 – 961.

Tvaryanas A.P., Thompson W.T., & Constable S.H. (2006). Human factors in remotely piloted aircraft operations: HFACS analysis of 221 mishaps over 10 years. Aviat Space Environ Med 77:724-32.

Tvaryanas A.P. (2006). Human systems integration in remotely piloted aircraft operations. Aviat Space Environ Med 77:1278-82.

Wickens, C. D. & Hollands, J. G. (2000): Engineering Psychology and Human Performance. Prentice Hall.

Wiener, E.L. & Nagel, D.C. (1988) Human Factors in Aviation. Academic Press San Diego CA Williams, K.W. (2004). A Summary of Unmanned Aircraft Accident/Incident Data: Human Factors Implications. Federal Aviation Administration, Department of Transportation Report No. DOT/FAA/AM-04/24.

Control of Multiple Unmanned Vehicles: a capacity model from a meta-analysis

Christopher D. Wickens, Mala Gosakan, Marc Gacy, Lila Laux

Alion Science and Technology, Boulder, Colorado

How many unmanned systems can the operator effectively control and/or supervise before the operator begins to experience higher mental workload, thus losing SA and degrading overall mission performance? In this paper we describe the results of a literature search that was undertaken to identify factors of particular importance to the control of unmanned systems. We present the architecture for a computational model of the controller of multiple UVs that accounts for the diminishing gains in overall mission productivity and eventual loss in productivity that appears to occur as the number of vehicles under supervision, N, increases.

A question asked with increasing frequency in advanced aviation is "how many unmanned air vehicles" can a human supervise? The challenge here is that as the number of unmanned air vehicles (UVs) grows, a single human operator who must both supervise an overall "fleet" and service each individual vehicle when required becomes taxed to the limit of his or her capabilities and beyond, thus diminishing the overall performance of the fleet. Our task in this research was to undertake a review of the literature on the supervision of UVs to identify both computational models and empirical data that could help to address this question, as well as those data that would identify mitigating factors to such a limit; particularly the degree of automation or autonomy provided per vehicle, (Wickens Dixon & Chang, 2003, Dixon Wickens & Chang, 2005) or overall automation decision support (Nehme et al, 2008).

As part of a larger literature review (See Appendix A of Gosakan, 2011 for more details), the goal of the research we report here was to identify empirical studies that had varied the number of USs under supervision, N, and provided objective outcome measures of overall mission performance. Our goal was to synthesize these results into a plausible model of the human controller of MUVs. Many of the concepts underlying this model are attributed to the extensive empirical research carried out by Cummings and her laboratory at MIT (e.g., Domneez et al, 2010, Cummings et al, 2010; Nehme et al, 2008); along with a smaller amount of research carried out by the first author at the University of Illinois (e.g., Wickens Dixon & Chang, 2003; Dixon Wickens & Chang, 2005, Wickens Levinthal & Rice, 2009, Wickens Dixon & Ambinder, 2008).

Results of Meta Analysis

The graph in Figure 1 presents an overview of the data from studies that had varied N and assessed some global measure of mission performance; such as the percentage of territory surveyed within a fixed amount of time, the number of targets destroyed, or the number of enemy discovered. The individual studies are identified by number in the box, and these numbers can be cross referenced to the citation list at the end of this paper. In the citation list, the number identification of each study (seen in figure 1) is boldfaced at the end of the citation. The performance metric on the Y axis is normalized to the maximum possible value within each study. The most striking aspects of this figure are the general non-linear increase in global mission performance with N, and the fact that some studies actually show an inverted U shaped pattern, suggesting that mission success has an optimum N, and that adding further UAVs (referred to as assets in sections below) to the skies actually diminishes overall performance.

An Asset-centric Model

In order to understand the causes of the general form of the data shown in figure 1, we present a model of the influences on mission performance as a function of N. In doing so, we distinguish between two important metrics: a **global performance** (GP) metric such as that depicted on the Y axis of figure 1, and a unit **asset productivity** (AP) metric, that characterizes the productivity of each individual asset. As N grows, global performance is generally expected to grow (with the exception of the high level down turn, accounted for by our model), but AP may not follow this pattern. In the following sections, we distinguish four different influences on AP, as they may be mediated by both N and by AP. The model we present first is one that is "asset-centric" in that it focuses on the performance of specific assets, to create the global performance metric. We then turn to defining an operator-centric model, based in part upon components of the asset model.

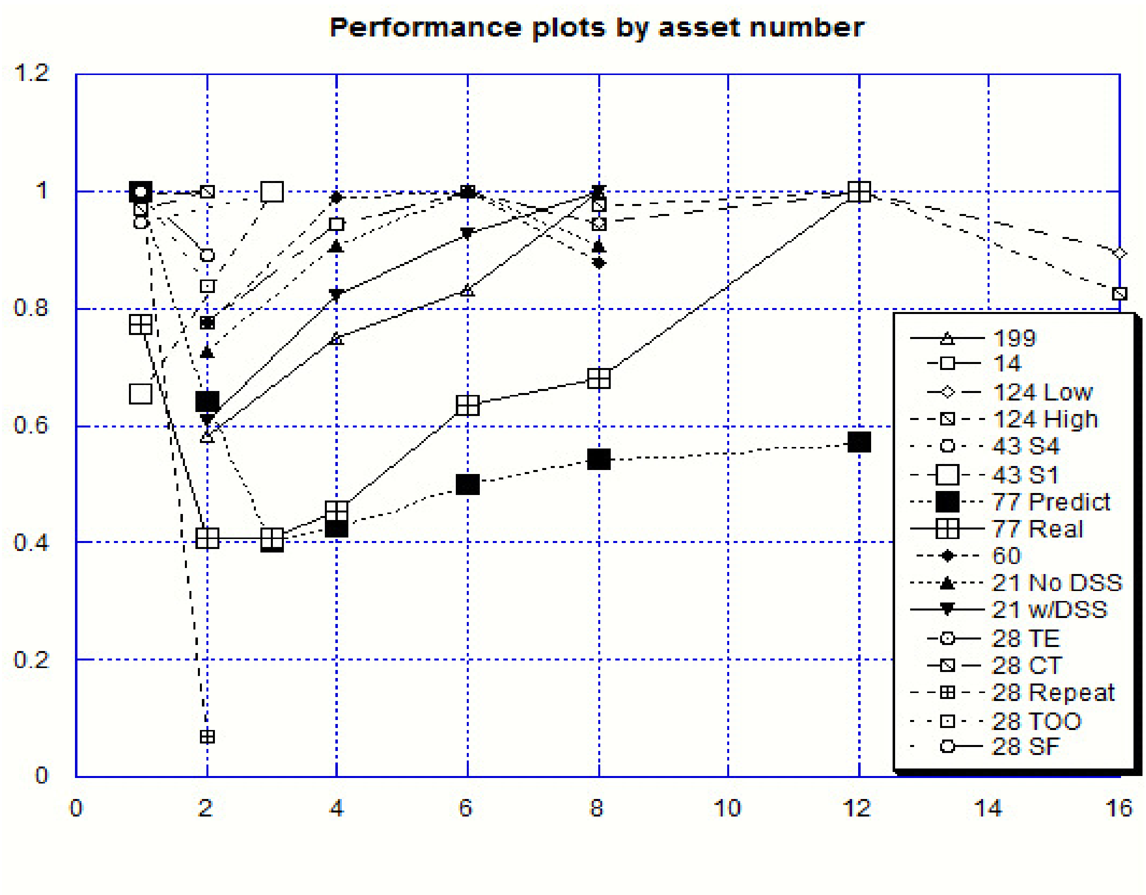


Figure 1**:** Global mission performance as a function of N (number of UVs).

First, as N increases this increases the total **power** of the fleet in a relatively linear function. For example as more surveillance vehicles are placed in the air, careful coordination will assure linearly increasing amounts of terrain covered, per unit time. Indeed some of the studies in Figure 1 appear to reflect this relationship (.e.g., study 199**).**

Second, this increase is not necessarily linear, as a pure fleet power function would predict, but often logarithmic in its approach to maximum. To account for this, we partition individual asset performance (AP) time into three separate components, following Nehme et al (2008) and Cummings & Mitchell, 2007 (see figure 2).

1. Service time (ST) is the amount of time it takes each asset to be "serviced" by the human supervisor; that is for example, directed on its next trajectory, or performing some maintenance function upon it. Assuming the operator can only service one asset at a time, an observation consistent with empirical data (Dixon & Wickens, 2003), then time servicing one asset must preclude servicing any other asset simultaneously.
2. Productive time (PT). This is time when the asset can operate autonomously in a way that is productive for the overall mission (e.g., capturing ground video along a designated track), and hence contribute to GP.
3. Wait time (WT). This is non-productive time when the asset must wait until servicing becomes available (Nehme et al, 2008). One can then define asset productivity as the ratio of PT to [PT +ST + WT). This ratio can also be defined as an **asset percent utilization.**

Given this ratio, and the single server nature of the queue, it then follows that as N increases, WT will increase linearly as NxWT, as a given asset must wait longer while a given number (N-1) of assets must be serviced. Hence the percent utilization of each asset will decrease, predicting the logarithmic increase in GP generally shown by the left side of the functions in Figure 1. We can see this influence of N near the top of Figure 2.

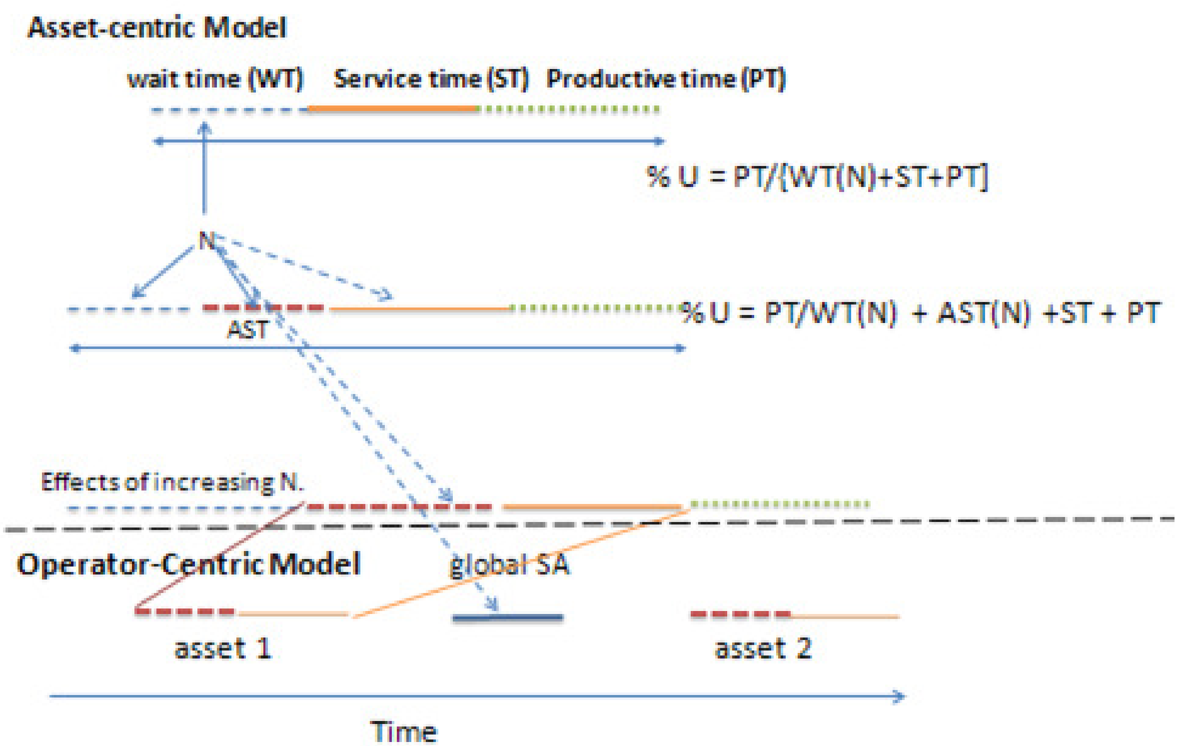


Figure 2. Components of the time cycle for each asset. The influence of N on these components is shown by comparing the second and third time line, where it can be seen that in the third time line the percent Utilization (productive time to total time) is diminished. On the bottom the foundations of the operator-centric model are depicted, showing the servicing of two assets and the time required to maintain global SA (the solid line).

A **third** influence on this timing cycle is shown in the middle of Figure 2. Here **attention switch time** (AST) is added as the time between when an asset appears for servicing, and when the supervisor can turn his full attention to servicing it. (Wickens Dixon & Chang, 2003). In the time line in Figure 2, AST is assigned between the waiting time period, and the servicing time period, reflecting the fact that when a new asset is "noticed" it may take some time for the supervisor simply to orient to the asset ("which asset is that, and where is it located?"), as well as to restore his or her memory for the mission of that particular asset, and how that mission should be altered (e.g, a change in trajectory, altitude, etc.). Importantly, we see that both aspects of this cognitive component will be influenced by N. As more units are present, the supervisor will spend increasingly more time identifying which unit is in need of servicing (or which is the unit requesting servicing), as well as choosing the nature of the service (e.g., the next mission segment). This choice will take longer to the extent that it has been longer since a given asset was serviced, because memory for what it was doing before, will have declined. And this time of operator neglect of a particular asset will grow with N. Both the nature of the increase in AST with N, and its slope remain unclear. As a default, it may be considered to be linear with N.

Within the model described so far, depending on the sub-function ***AST = f(N)***, the shape of the global performance curve in Figure 1 may reach a plateau sooner or later (e.g., at a lower or higher value of N), but also has the possibility of inverting to a downward slope at the right side, as shown by some of the functions in Figure 1. This will happen if the penalties imposed on WT and AST more than offset the gains in global productivity (GP) achieved by increasing N. However there is a **fourth** mechanism that can make this downward slope on the right side of Figure 1 even more pronounced, and this describes what may be called "**global situation awareness**"; that is, maintaining the "big picture" of who is where and how the mission is progressing. It is well known that situation awareness depends in part upon working memory (Endsley, 1995; Wickens, 2002, 2008), and that working memory demands will be heavily influenced by the number of entities which must be "kept track" of (e.g., Hess, Detweiller, & Ellis 1999). This increase in working memory demand can have two consequences: (a) to the extent that the supervisor makes some effort to sustain working memory during servicing of each item, this may impose a penalty on servicing time (shown by the dashed light arrow in figure 2). and (b) the added demands on working memory may diminish overall situation awareness (keeping track of more things is harder), and hence the overall quality of planning (e.g., optimal queuing and service order of assets; optimal assignment of assets to new tasks) will degrade, leading to an effect not on time, but on the accuracy of global performance, the Y axis of Figure 1).

In the above representation, it is important to note that, just as N will decrease asset productivity via several components in the equation, so anything that influences productivity via automation will increase this productivity. For example longer autonomous trajectories, or more artificial intelligence within the asset will reduce the frequency of required servicing and hence increase the asset productivity or utilization ratio depicted in the middle row of Figure 2. Furthermore, the presence of decision support tools to assist the supervisor during the switching period can substantially reduce this time penalty, as well as improve the overall accuracy of performance. Indeed, one study (Nehme et al, 2008) compared the presence and absence of decision support tools, and observed the inverted U pattern when such tools were absent, but a linearly increasing pattern when they were present to assist the operator in servicing choices (see study **21** in figure 1)..

An Operator-Centric Model

The top three rows of Figure 2 depicted a time line for the cycle of use of a single asset – the asset-centric model. Within this figure, the middle two components, ST, & AST were directly associated with human operator time, the first of these also being affected by N. At the bottom of figure 2 is depicted the foundations of the operator centric model. Here we have extracted the two components of human operator time from above and depicted (on a condensed time line) these as they might characterize the servicing of 2 assets by the human operator. Two features are of note here. First, in this figure, there is a good deal of unfilled (slack) time, in which the operator's resources are idle, or "underutilized". Second, part of this slack time is filled with a solid line time period, corresponding to the fourth mechanism described in the previous section; that is, the need to maintain **global SA** about how the mission is going, how the fleet is coordinated, and whether particular assets may need to be re-assigned to assist others.

Second, with the current demands depicted in the figure, there is ample time available to schedule this global SA task between arrivals of assets for servicing (these are the unfilled portions of the time line). However it should be apparent, that as either servicing time or AST increases (with increasing N), this free time will diminish, and eventually reach a point at which there is not enough time to develop and maintain global SA (or if it is given such time, then AST and ST will not be allocated sufficient time). Hence global performance will degrade, either because of the shortchanging of overall SA, or because adequate service and attention switch time is not available for each asset (or both) thus resulting in the operator committing errors.

According to the components of the model then, the composite inverted U shaped function revealed by the meta-analysis and portrayed in figure 1, may be seen to be made up of two additive components. One is the logarithmic function showing the increased mission power with N, but showing the diminishing gains with N, as wait time (and switching time) increases, decreasing individual asset productivity. The second component is the actual diminishing function with N, related particularly to the loss of SA, as other elements compete with this essential time demand. The precise peak of the function of course depends on the relative slopes of the two individual functions, as well as the possible mitigating role of automation (shifting that peak to the right); however the existing data suggest that it may be somewhere between 3 and 8. Continued research along these lines is necessary to better understand the quantitative role of the two critical human performance components in the model: attention switching time and SA maintenance, and the precise manner in which these are influenced by N.

Acknowledgments

The research reported in this document was performed in connection with contract LWI 400-145. The authors would like to acknowledge Sirjana Dahal at Missouri University of Science & Technology and Dr, Lila Laux and Patty Keesler at Alion Science & Technology for their contributions to this effort.

References:

Bold faced numbers identify studies depicted in figure 1.

Crandall, J., Cummings, M., Nehme, C. (2008). A Predictive Model for Human-Unmanned Vehicle Systems. In (HAL2008-05), MIT Humans and Automation Laboratory, Cambridge, MA. **{60}**

Cummings, M., Bruni, S., Mercier, S., Mitchell, P. (2007). Automation Architecture for Single Operator Multiple UAV Command and Control. In Decision Support for Network-Centric Command and Control 2 The International C2 Journal | Vol.1, No 2l. **{77}**

Cummings, M., Bruni, S. & Mitchell, P. (2010) Human supervisory control challenges in network-centric operations., *Reviews of Human Factors & ergonomics, 6*, 34-78.

Cummings, M., Guerlain, S. (2007). Developing operator capacity estimates for supervisory control of autonomous vehicles. In Human Factors: The Journal of the Human Factors and Ergonomics Society February 2007, Vol. 49, no. 1, pg 1-15.**{124}**

Cummings, M., Mitchell, P. (2007). Predicting Controller Capacity in Supervisory Control of Multiple UAVs. In Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, Vol.38, Issue 2, pg 451-460.

Dixon, S.R., Wickens, C.D., & Chang, D. (2005). Mission control of multiple UAVs: A quantitative workload analysis. Human Factors, 47, 479-487 **{14}**

Donath, D., Rauschert, A., Schulte, A. (2010). Cognitive Assistant System Concept for Multi-UAV Guidance Using Human Operator Behaviour Models. In Second International Conference on Humans Operating Unmanned Systems. HUMOUS'10. **{43}**

Donmez, B.D., Nehme, C., & M.L. Cummings. (2010). Modeling Workload Impact in Multiple Unmanned Vehicle Supervisory Control. In IEEE Systems, Man, and Cybernetics, Part A Systems and Humans, Vol. 99, p. 1-11.

Endsley, M. R. (1995) Towards a Theory of Situation Awareness in Dynamic Systems, *Human Factors, 37*, 32-64.

Gosakan, M. (2011). HSI Tool Enhancements to Better Represent Operator Workload When Controlling Unmanned Systems. Final Report submitted to Leonard Wood Institute.

Hess, S. M., Detweiler, M. C., & Ellis, R. D. (1999). The utility of display space in keeping-track of rapidly changing information. Human Factors, 41, 257-281.

Nehme, C., Crandall, J., Cummings, M. (2008). Using Discrete-Event Simulation to Model Situational Awareness of Unmanned-Vehicle Operators. In Virginia Modeling, Analysis and Simulation Center Capstone Conference, Norfolk, VA, April 2008.

Nehme, C., Mekdeci, B., Crandall, J., Cummings, M. (2008). The Impact of Heterogeneity on Operator Performance in Future Unmanned Vehicle Systems. In The International Command and Control Journal, Vol. 2(2), 2008. **{21}**

Pina, P., Cummings, M., Crandall, J., Della Penna, M. (2008). Identifying Generalizable Metric Classes to Evaluate Human-Robot Teams. In Metrics for Human-Robot Interaction Workshop at the 3rd Annual Conference on Human-Robot Interaction, Amsterdam, Nederland. **{199}**

Wickens, C.D. (2002). Situation awareness and workload in aviation. Current Directions in Psychological Science, 11(4), 128-133

200) Wickens, C.D. (2008). Situation awareness. Review of Mica Endsley's articles on situation awareness. Human Factors, Golden Anniversary Special Issue, 50, 397-403.

Wickens, C., Dixon, S., Ambinder, M. (2006). Workload and Automation Reliability in Unmanned Air Vehicles. In N. J. Cooke, H. Pringle, H. Pedersen, & O. Connor (Eds.), Advances in Human Performance and Cognitive Engineering Research, Vol. 7, Human Factors of Remotely Operated Vehicles (pp. 209-222). Elsevier Ltd.

Wickens, C., Dixon, S., Chang, D. (2003). Using Interference Models to Predict Performance in a Multiple-Task UAV Environment - 2 UAVs. In Technical Report, April 2003 **{28}**.

Wickens, C., Levinthal, B., & Rice, S. (2009). Imperfect Reliability in Unmanned air Vehicle Supervision and Control. In M. J. Barnes & F. G. Jentsch (Eds.), Human-Robot Interaction in Future Military Applications (pp. ??) Hampshire, England: Ashgate Publishing.

DEVELOPMENT AND VALIDATION OF MEASURES FOR ARMY AVIATION COLLECTIVE TRAINING

Martin Bink
U.S. Army Research Institute
Fort Benning, GA
Melinda Seibert
Courtney Dean
Aptima, Inc.
Woburn, MA
John Stewart
U.S. Army Research Institute
Fort Rucker, AL
Troy Zeidman
Imprimis, Inc.
Huntsville, AL

Simulation-based Aviation Training Exercises (ATX) are critical for preparing U.S. Army Combat Aviation Brigades for deployment. However, while offering the opportunity to practice mission segments at the unit level, the effectiveness of this training remains unclear due to a need for objective assessments focused on observable team behavior. Unit Commanders and trainers need tools for measuring collective task performance in order to understand performance gains, facilitate feedback, and guide the learning of aviation tactical teams. To address this challenge, a set of aviation team performance measures were developed, data were collected to validate these measures, and strategies were created to facilitate application of the measures to collective training events. The measures used behaviorally-based observations to assess performance of aviation tactical teams. The measures were used at multiple ATX events to assess performance of aviation tactical teams. Data were collected on inter-rater reliability and on agreement between the measures and overall mission performance. Results provided evidence of both acceptable reliability and validity for the measures. Moreover, requirements were developed for electronic data collection tools that can be used by unit Commanders and trainers to assess team performance at collective training exercises.

Previously, unit-level collective aviation training was accomplished through live field exercises. However, for many reasons (e.g., limited resources and lack of access to suitable practice areas), live training is less feasible than in the past. A response to these limitations was the development of the U. S. Army Aviation Warfighting Simulation Center (AWSC), a networked training system located at Fort Rucker, Alabama. The AWSC consists of 24 networked cockpit simulators that can be reconfigured to represent the Army's four currently operational combat helicopters (AH-64D Apache, CH-47D/F Chinook, OH-58D Kiowa Warrior, and UH-60 A/L/M Blackhawk). Using the AWSC, a Combat Aviation Brigade (CAB) can participate in a collective Aviation Training Exercise (ATX) that places CAB aircrews and battlestaff in a common virtual environment. While the primary purpose of ATX is to assess the readiness of battlestaff, ATX also provides an opportunity for feedback on the readiness of aircrews. The challenge addressed here is to develop methods to facilitate the provision of feedback on collective skills and task performance in a manner that meaningfully guides further development at the aviation tactical level (e.g., Company and below)

Even though individual aviation tasks are generally well defined, aviation collective tasks are comparatively poorly defined as broad mission segments that Army Aviation teams must accomplish (Cross, Dohme, & Howse, 1998). Army aviation collective tasks for reconnaissance and attack operations refer to those aviation tasks that require coordination between one aircraft and another, coordination between an aircraft (or flight of two or more aircraft) and a tactical command element (e.g., Brigade Aviation Element), and coordination between an aircraft and a Ground Commander. While tools exist to help aviators obtain step by step lists of actions to be performed, requisite underlying knowledge and skills that support aviation collective tasks cannot be inferred from such broad functions within those tasks or from task descriptions alone which lack objective performance criteria. Rather, behaviorally-anchored indicators of aviation team performance, which link observable behaviors to discrete benchmarks, should be used to evaluate performance on aviation collective tasks.

Training research (e.g., Salas, Rosen, Burke, Nicholson, & Howse, 2007; Salas, Rosen, Held, & Weissmuller, 2009; Stewart, Dohme, & Nullmeyer, 2002; Stewart, Johnson & Howse, 2007) has demonstrated that the lack of clear performance assessment criteria fails to fully exploit the effectiveness of simulation-based training events. Moreover, the military value of simulation-based training, such as ATX, is determined by performance improvement of participants within the virtual-training environment (Bell & Waag, 1998). In the case of ATX, there is a need to develop performance criteria on aviation collective tasks in order to clearly illustrate what right looks like for aircrews and leaders and to assist Observer-Controllers (OCs) in providing feedback.

The primary objective of this research effort was to develop a reliable, valid, and useful assessment system. Using this system, unit leaders and OCs could provide consistent behaviorally-based feedback to aircrews that would help distinguish high-performing teams from low-performing teams. Performance results from across training units could then be aggregated to provide unit leadership with a “snap shot” of proficiency on aviation collective tasks, resulting ultimately in better performing teams. To achieve this objective, observer-based measures of aviation performance in mission-critical collective tasks were first defined. The measures were then implemented into a hand-held electronic tablet and OCs and unit leaders rated aviation team performance in multiple ATXs. Reliability and validity analyses were conducted to identify whether the observer-based measures accurately, consistently, and appropriately predicted team performance.

Measure Development

The measures were constructed using the Competency-based Measures for Performance ASsessment Systems (COMPASS[SM], MacMillan, Entin, Morley, & Bennett Jr., in press) approach. COMPASS is a methodology for the development of performance measures that combines experiential knowledge of subject matter experts (SMEs) with established psychometric practices. A set of three SME-based workshops took place over the course of five months that moved from the identification of key observable behaviors to the construction of performance measures. The first and third workshops were group interviews while the second workshop consisted of individual or small group interviews. A total of 27 SMEs participated across all workshops, including 3 SMEs participated in all three workshops. SME expertise ranged from military aviators to simulation training experts and software engineers.

In the first step of measure development, the phases of the attack/reconnaissance mission were deconstructed into observable behaviors, or performance indicators (PIs), that allow an expert to recognize whether an individual or team is performing well or poorly. The resulting PIs and relevant missions/tasks provided a solid basis on which to develop benchmarked measures that are less sensitive to subjective biases and more reliable over repeated sessions. In the second step, SME-provided information was crafted into specific performance measures associated with each PI in order to create performance measures with appropriate behaviorally-based rating scales (i.e., 5-point Likert-type scales). To obtain exemplar behavior information, SMEs were asked to describe and identify explicit behaviors that were representative of good, average, and poor performance. Throughout the measure development process, care was taken to ensure that measures were operationally relevant, thorough, and appropriately worded using domain language and terminology. Altogether, 130 candidate observer-based performance measures were developed. Table 1 provides an example for the PI *Request Clearance of Fires from Ground Commander*. In the final step, SMEs were presented the full set of measures to review and revise as required to ensure the measures could be understood and accepted by a wide range of potential users. Modifications were made to the measures, resulting in a final list of 115 performance measures for assessing the performance of an aviation collective team performing an attack/reconnaissance mission.

Table 1.
Example Performance Measure - Request Clearance of Fires from Ground Commander.

Does the flight request clearance of fires from Ground Commander?				
1	2	3	4	5
Flight does not request clearance of fires		Flight considers ROE; establishes friendly/enemy positions; requests clearance of fires; not ready to effect the target while going through this process		Flight considers ROE; establishes friendly/enemy positions; requests clearance of fires; anticipates clearance and sets up shot during this process

Measures Reliability and Validity

Inter-rater reliability was first evaluated as the intended use of the measures requires that different raters use the scale similarly. After demonstrating acceptable reliability, criterion-related validity was explored to determine if measures relate to performance outcomes in aviation tactical missions. The ultimate goal of reliability and validity analyses in this effort was to evaluate how well measures performed and to inform revisions to the measures and scale anchors as appropriate.

Method

Reliability and validity data were obtained during two separate ATX events conducted at Fort Rucker, AL. A total of 21 missions across two different units were observed. Of the 21 missions, 15 were simultaneously rated by two or more experienced aviators. Three of those 15 featured three independent raters. The remaining six missions were rated by one experienced aviator. Outcome measures were obtained from 21 missions and focused on more objective outcomes of the mission (e.g., mission accomplishment, achievement of objectives, number of targets destroyed, aircraft lost). While raters evaluated flight team performance in real-time, outcomes measures were completed following the end of a mission, both collected using an electronic measurement tool. Given these data, inter-rater reliability was evaluated on the 15 missions with multiple raters in each while criterion-related validity was examined on all 21 missions.

In the absence of an existing pure criterion measure (i.e., an independent objective training or performance outcome) a substitute measure was developed. This outcome measure consisted of nine items indicating variables such as mission success, number of targets destroyed, number of friendly aircraft lost, and instances of fratricide. Given limited access to higher-level-leader raters, outcome ratings were completed by the same observers who rated the process measures. While this analysis does not speak directly to criterion validity because of rater dependencies and the absence of a true criterion, it serves as way to verify consistency and relationships between process and outcome measures.

Results

Inter-rater reliability. While inter-rater reliability is a standard approach for demonstrating that raters use measures and scale anchors similarly, evaluations of other measure properties such as percent agreement can be insightful tests of the reliability of ratings (Howell, 1997). Further, percent agreement as computed in this study can help identify measures that were especially problematic for raters to agree upon – an important step for revising as well as down-selecting the large measures set to a manageable number of the best performing and useful items. As a result, inter-rater agreement was first assessed and then followed up with a more standard inter-rater reliability analysis.

Inter-rater agreement was established using a percent agreement method based on the range of ratings for each measure across the raters (e.g., both raters within one rating point). For each level of agreement, percent agreement was calculated by dividing the observed agreement counts by the total number of possible observations. When aggregated across all rated missions, raters achieved a 72% agreement within 1-point on the Likert scales. Put differently, if one rater gave a rating of five, the other rater(s) was likely to give a rating of at least four in 72% of the occasions. Considering the many uncontrollable environmental factors present during this testing, these results are quite promising in demonstrating that raters would use scale anchors similarly.

Given the relatively high percent agreement observed in the first analysis, inter-rater reliability was computed using Cohen's Kappa (κ), a conservative measure of inter-rater agreement that accounts for chance agreement (Cohen, 1960; Fleiss, 1981). Reliability was substantial ($\kappa = 0.66$) with the 1-point-agreement threshold. Overall, the analyses suggested that different raters similarly interpreted the collective task measures. However, these results also suggested that some measures were not achieving high levels of reliability. Given these initial findings, along with the goal of refining the measures, further examination assessed which specific measures tended to have lower and higher levels of agreement.

Criterion-related validity. Only the most reliable measures were included in this analysis (i.e., rating agreement at or within 1-point in 80% of the observations). Performance measure averages were computed for each mission and were compared to average ratings for corresponding outcome measures. There was a positive relationship between performance and outcome measures such that higher ratings on performance measures were associated with higher outcome scores ($r = 0.48$, $n = 32$, $p < 0.05$). This result suggested that the performance measures developed to assess Army aviation collective skills do predict performance outcomes and are, therefore useful and valid predictors of performance. Taken as a whole, the results suggest that while the developed measures have some degree of validity, further work is required to refine the whole measures set prior to full implementation. Combined with reliability data, these results offer guidance on how to revise and construct the most effective measure set.

Discussion

The primary objective of this research effort was to develop reliable, valid, and useful tools to assist Leaders and trainers in assessing aviation collective performance. Using these measures, trainers are anticipated to be better able to provide consistent, behaviorally-based feedback that can help to improve the performance of aviation teams. Here, the focus was on collective tasks critical to performing typical scout/reconnaissance missions. More generally, beyond ATX, these measurement tools could also be useful in preparing for and conducting assessments in a variety of Army aviation collective training events (e.g., at home station).

The research effort reported here resulted in the construction of 115 draft measures focusing on key skills for flight teams in collective tasks. For these draft measures, initial data concerning reliability and validity were collected. These data provided evidence that the measures are in general reliable, and suggested a modest correlation between reliable performance measures and outcome measures. It should be noted that while the findings on reliability and validity were limited and preliminary, these analyses provided data on the subsets of measures that are most and least reliable, which enabled measure revision and refinement. In addition, information was collected on the requirements for tools to best enable use of the measures that will guide subsequent implementation. Based on these findings, the measures set was reduced to 105 well-performing measures, and strategies to facilitate their use were identified.

Collectively, these findings support a scientifically-based implementation plan that is designed to create a comprehensive performance assessment system. Several key objectives of this plan are to:

- Implement the refined observer-based performance measures in hand-held, tablet-based measurement tools to enable organization of measures and electronic capture of ratings for debriefing and performance tracking.
- Explore and implement related system-based measures that, once combined with observer-based metrics, could enable a more complete assessment of collective skills through leveraging of simulator data streams.
- Design and create debriefing tools that provide targeted feedback on team performance.

Ultimately, these measurement tools will enable OCs to evaluate aviation teams as they perform collective tasks at ATX. Similar evaluations by unit leaders and instructor pilots are anticipated to be possible using these tools in other collective training environments as well. Such evaluation can illuminate the status of underlying knowledge and skills and enable formative feedback that is likely to guide learning and foster development of strong teams in collective training events.

References

Bell, H. H., & Waag, W. L. (1998). Evaluating the effectiveness of flight simulators for training combat skills: A review. *International Journal of Aviation Psychology, 8*, 223-242.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37-46.

Cross, K.D., Dohme, J.A., & Howse, W.R. (1998). *Observations about defining collective training requirements: A White Paper prepared in support of the ARMS program*. (ARI Technical Report 1075). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA349437).

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York: Wiley, (2nd ed).

Howell, D. C. (1997). *Statistical methods for psychology (4th ed.). Belmont, CA: Duxbury Press.*

MacMillan, J., Entin, E. B., Morley, R. M., & Bennett Jr., W. R. J. (in press). Measuring team performance and complex and dynamic military environments: The SPOTLITE method. *Military Psychology.*

Salas, E., Rosen, M.A., Burke, C.S., Nicholson, D., & Howse, W.R. (2007). Markers for enhancing team cognition in complex environments: The power of team performance diagnosis. *Aviation, Space, & Environmental Medicine, 78*, B77-85.

Salas, E., Rosen, M.A., Held, J.D., & Weilssmuller, J.J. (2009). Performance measurement in simulation-based training. *Simulation & Gaming, 40*, 328-376.

Stewart, J. E., Dohme, J. A., & Nullmeyer, R. T. (2002). U.S. Army initial entry rotary-wing transfer of training research. International Journal of *Aviation Psychology, 12*, 359-375.

Stewart, J. E., Johnson, D. M., & Howse, W. R. (2007). *Fidelity requirements for Army aviation training devices: Issues and answers*. (ARI Research Report 1887). Arlington, VA: U. S.

EXPLORING THE BOUNDARIES OF COMMAND AND CONTROL MODELS OF DISTRIBUTED TEAM PERFORMANCE IN AVIATION AND AEROSPACE OPERATIONS

Haydee M. Cuevas
Embry-Riddle Aeronautical University
Daytona Beach, FL

Barrett S. Caldwell
Purdue University
West Lafayette, IN

Traditional command and control (C2) models focus on a centralized command managing and directing one or more subordinate elements to perform required functions. However, in distributed C2 environments, a human commander has less ability to fully understand and control the behavior of "agents" (either human domain experts or autonomous automated systems) in real-time operations. In this paper, we explore the situational, information, and human performance issues that constrain the appropriateness of classical C2 system design, and highlight the need for distributed C2 information flow capabilities, in contemporary human-human and human-automation teams. We discuss these issues in the context of modern day aviation and aerospace operations.

Both organizational command and control (C2) structures and human supervisory control paradigms (including human-automation interaction architectures) are based on models of command information flow and agent behaviors in complex systems. C2 involves the functions of planning, directing, coordinating, and controlling operations to achieve organizational objectives (U. S. Department of Defense, 1994). Traditional C2 models focus on a centralized command managing and directing one or more subordinate elements to perform these functions. However, in distributed C2 environments, a human commander has less ability to fully understand and control the behavior of "agents" (either human domain experts conducting coordination tasks, or autonomous systems with increasing levels of on-board capabilities) in real-time operations. Thus, as modern systems continue to grow in complexity and scope, a greater need has emerged for C2 models that flexibly adapt to task and situational constraints. Accordingly, in this paper, we explore the situational, information, and human performance issues that constrain the appropriateness of classical C2 system design, and highlight the need for distributed C2 information flow capabilities, in contemporary human-human and human-automation teams. We discuss these issues in the context of modern day aviation and aerospace operations (e.g., unmanned aerial systems, long duration spaceflight missions).

Command and Control of Distributed Expertise

Distributed team members share knowledge and understanding of the world based on varying levels of expertise in a variety of specialized domains, interactions with distinct or overlapping system components, and availability of shared as well as individual information (Salas, Cannon-Bowers, & Johnston, 1997). These distinct capabilities can be thought of as multiple dimensions of expertise, and not simply different levels of expertise in a single subject matter area (Caldwell, 2005). However, distributed expertise may alter coordination within the C2 structure in that the ‘commanded’ agents may have a different understanding of the situation as well as unique capabilities for processing current information and integrating system status information. Such agents may be able to develop independent or superior awareness, execution parameters, or updated evaluations of task requirements that might make obsolete a commander's orders. This problem is exacerbated when dealing with large teams consisting of teams of teams, as often found in military network-centric operations (e.g., Gorman, Cooke, & Winner, 2006).

Of particular interest in such domains is *communication efficiency*. This concept reflects a process of determining when meanings are shared or not, and then being more explicit in a reference to a term with differing meanings across members of a functioning team. Unfortunately, in distributed teams, such unshared meanings may not be visible, as each team member may be clear on their own meaning, and unaware that others also have a clear understanding--but not of the same meaning (e.g., "positive feedback" has a negative connotation for a controls engineer, but a positive connotation for an industrial / organizational psychologist) (cf. Qureshi & Vogel, 1999). In aviation, dynamic, fluid teams benefit from a strong, shared, and generalized professional, operational, and cultural training process, and thus, this would be expected to facilitate communication efficiency. For example, flight crew can refer to the same shared training processes and patterns of implicit meanings from working at the same airline with the same organizational culture and training programs. However, in spaceflight operations, involving individuals with a broad and diverse range of training, specialization, experience, and cultural background, the challenge becomes how to enable a distributed team of experts to perform successfully as a distributed expert team (cf. Stagl et al., 2007).

As the complexity of both the operational scenario and the organizational architecture increase, the classical expectations of a commander capable of directing and anticipating the needs and instructional demands of each commanded agent become less viable. For instance, the structure of a spaceflight mission control team consists of multiple flight controllers, each with his or her own technical subsystem domain. This group of technical subsystem controllers is integrated by controllers with responsibilities for shared displays and computer systems (translating incoming data to synchronized information presentation) as well as Flight Directors responsible for coordinating the technical subsystems to achieve strategic goals (translating information to shared knowledge) (Caldwell, 2000; 2005). This model represents a coherent functional architecture that allows for coordination of multiple domains of expertise, any one of which may achieve greater detailed local awareness than the supervisory Flight Director. As will be discussed in the next section, C2 of distributed expertise is further complicated when the commanded agents are both human and non-human entities.

Information and Communication Technology in C2 Operations

The integration of information and communication technology (ICT) into C2, as evident in aviation, aerospace, and military network-centric operations, has resulted in the increased prevalence of human-automation teams, comprised of both human experts and expert automated systems. This creates unique challenges that must be overcome to ensure effective coordination among distributed members. A socio-cognitive perspective emphasizes that human operators’ beliefs about and trust in automation dictate their subsequent reliance on automated systems (such as intelligent agents or decision support systems), ranging from the extremes of over-reliance and complacency to under-reliance and mistrust (Cuevas, Strater, Caldwell, Fiore, & 2007; Lee & See, 2004). Fortunately, this issue can be addressed through adequate training regarding the automation's functional capabilities (e.g., reliability) and limitations (e.g., effects of contextual factors) as well as through appropriate system display design (i.e., information presentation in terms of content and format) (Lee & See, 2004). Furthermore, true 'collaboration' in distributed human-automation teams is somewhat at odds with classical C2 models. In modern human-automation teams, the automated system (e.g., intelligent agent) can be viewed as a true partner collaborating with human experts and not simply a directed entity or useful collaboration tool (Cuevas, Fiore, Caldwell, & Strater, 2007). The integration of ICT in human-automation teams, therefore, is enabling a type of flattening or upward flow of information that would not be possible in traditional C2 architectures.

To illustrate, ICT, such as web-based C2 systems, supports military operations by allowing commanders to engage subordinate leaders and staffs in collaborative planning and decision making at all levels within units (Riley, Endsley, Bolstad, & Cuevas, 2006; U. S. Army, 2001). By supporting these critical processes, this technology enables commanders to reduce decision cycles within their organizations. Web-based C2 systems also facilitate the rapid dissemination of orders based on the commander's decisions to the lowest levels, thus maximizing time available for tactical units to prepare for, synchronize, and initiate decisive action. However, the limitations associated with the ICT used may significantly influence effective communication and coordination among distributed team members. For example, studies on computer supported collaborative work and groupware have highlighted both the benefits and potential drawbacks of introducing new ICT into distributed team interactions (e.g., Nunamaker, 1997; Olesen & Myers, 1999; Qureshi & Vogel, 1999). In particular, one perspective is to categorize human actors in terms of either users of ICT or those that structure the ICT for the users in the process of technology-use mediation (Orlikowski et al., as cited in Olesen & Myers, 1999). Structuring the ICT involves adapting the new technology within the context of use to facilitate integration into the organization as well as modifying the context, as appropriate, to accommodate the use of the new technology. Thus, changes in technology may force teams to restructure the patterns of information flow among members as well as the nature of their work (Mcgrath, Arrow, Gruenfeld, Hollingshead, & O'Connor, 1993).

In aviation and aerospace operations, the ICT available to distributed team members may also differentially influence team interactions and information flow. As an example, interest in using unmanned aerial systems (UAS) for a broad range of purposes is increasing at an unprecedented pace, making integration of UAS into the National Airspace System a priority for the Federal Aviation Administration. Modern day UAS include significant technological advances, blending automation with dynamic, decentralized control. In particular, UAS have fundamentally different aircraft control and communication architectures from manned aircraft due to the remote pilot location. Thus, a critical challenge facing the UAS community is to develop ICT that enables operators to interactively manage the flow of information from UAS with varying levels of autonomy while also facilitating collaboration with other UAS teams and support personnel. One area of particular interest is developing, testing, and fielding a C2 architecture for increasing multiple UAS control capability. Supporting this capability will require technology that facilitates seamless transitions between automation levels, situation assessment aids, and distributed teaming (Fern et al., 2011).

Within the context of spaceflight, given the task demands and system dynamics of NASA Mission Control Center operations, it is impossible for a unitary centralized C2 structure to effectively perform the required tasks to achieve mission success (Caldwell, 2000). Multiple ground personnel (e.g., mission control operators, research scientists, and other space mission support personnel) must communicate information and coordinate their efforts, not only among each other but with the astronaut crewmembers in space. Further, technological advances that have made possible longer duration manned missions, such as human exploration of Mars, have also brought about a concomitant push to reevaluate the concept of crew autonomy versus ground control (Kanas, 2005). For example, for long duration spaceflight requiring higher levels of crew autonomy, the role of the Mission Control Center may transition to that of Mission Support Center; Earth-bound ground personnel would take part in strategic and tactical mission planning and mission evaluation activities, and planetary explorers would be responsible for execution-level mission planning and mission execution (Grant et al 2006). In this domain, a more decentralized C2 architecture would be more practical and personnel require ICT that can enable these decentralized operations. Thus, ICT must effectively support the range of information flow processes and coordination activities (e.g., information exchange; coordination of distributed expertise) required by personnel, both on board the spacecraft and in the Mission Control Center, to perform their tasks safely and efficiently (Caldwell, 2006).

Conclusion

As organizations continue to evolve and integrate even more advanced ICT capabilities, traditional models of C2 must similarly mature in order to flexibly adapt to the challenges faced by distributed expert teams comprised of both human experts and expert automated systems. A critical challenge is ensuring that human operators have an accurate and complete understanding of the capabilities and limitations of supporting information technologies (Cuevas, Fiore et al., 2007; Strater et al., 2011; Strater et al., 2012). Equally important is recognizing how the integration of advanced ICT into C2 operations alters information flow and management. For example, web-based collaborative tools may actually circumvent the ability to conduct a strict C2 chain of information flow by making more paths of communication open with less control by centralized commanders. Additionally, the shift to new information flow capabilities may require a concurrent shift and re-prioritization of tasking to match operators' level of knowledge and experience with such systems (Caldwell & Cuevas, 2008; Caldwell, Palmer, & Cuevas, 2008). The continued evolution of C2 organizational architectures to include increasingly sophisticated information systems requires identifying and addressing the effects these changes will have on technology-mediated information flow and team coordination.

Acknowledgements

The views herein are those of the authors and do not necessarily reflect those of the organizations with which the authors are affiliated. Address correspondence to Haydee M. Cuevas at cuevash1@erau.edu.

References

Caldwell, B. S. (2000). Information and communication technology needs for distributed communication and coordination during expedition-class space flight. *Aviation Space and Environmental Medicine, 71* (9, Supp), A6-10.

Caldwell, B. S. (2005). Analysis and modeling of information flow and distributed expertise in space-related operations. *Acta Astronautica, 56* (9-12), 996-1004.

Caldwell, B. S. (2006). Group performance and space flight teams. In C. A. Bowers, E. Salas & F. Jenstch (Eds.), *Creating high-tech teams: Practical guidance on work performance and technology* (pp. 161-182). Washington, DC: American Psychological Association.

Caldwell, B. S. & Cuevas, H. M. (2008). Coordinating cycles of activity in distributed expert teams: An "Information Clutch" metaphor. *Proceedings of the 9th International Symposium on Human Factors in Organizational Design and Management* (ODAM 2008) [CD-ROM], São Palo, Brazil, 19-21 March 2008.

Caldwell, B. S. Palmer, III, R. C., & Cuevas, H. M. (2008). Information alignment and task coordination in organizations: An 'information clutch' metaphor. *Information Systems Management, 25 (1)*, 33–44.

Cuevas, H. M. Fiore, S. M., Caldwell, B. S., & Strater, L. (2007). Augmenting team cognition in human-automation teams performing in complex operational environments. *Aviation, Space, and Environmental Medicine, 78* (5, Supp Section II), B63-70.

Cuevas, H. M., Strater, L., Caldwell, B. S., & Fiore, S. M. (2007). Team cognition in human-automation teams. In K. Mosier & U. Fischer (Eds.), *Proceedings of the Eighth International Conference on Naturalistic Decision Making* [CD-ROM], Pacific Grove, CA, 3-7 June 2007.

Fern, L., Shively, R. J., Draper, M, H., Cooke, N. J., & Miller, C. A. (2011). Human-automation challenges for the control of unmanned aerial systems. *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting* (pp. 424-428). Santa Monica, CA: Human Factors and Ergonomics Society.

Gorman, J. C., Cooke, N. J. & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics, 49* (12-13), 1312–1325.

Grant, T., Soler, A. O., Bos, A., Brauer, U., Neerincx, M., & Wolff, M. (2006). Space autonomy as migration of functionality: The Mars case. *Proceedings of the Second IEEE International Conference on Space Mission Challenges for Information Technology*, Pasadena, CA.

Kanas, N. (2005). Interpersonal issues in space: Shuttle/Mir and beyond. *Aviation, Space, and Environmental Medicine, 76(2)*, B126-B134.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46* (1), 50-80.

McGrath, J. E., Arrow, H., Gruenfeld, D. H., Hollingshead, A. B., & O'Connor, K. M. (1993). Groups, tasks, and technology: The effects of experience and change. *Small Group Research, 24* (3), 406-420.

Nunamaker, J. F. (1997). Future research in group support systems: Needs, some questions and possible directions. *International Journal of Human-Computer Studies, 47* (3), 357-385.

Olesen, K. & Myers, M. D. (1999). Trying to improve communication and collaboration with information technology: An action research project which failed. *Information Technology and People, 12* (4), 317-332.

Qureshi, S. & Vogel, D. (1999). Organisational challenges and research directions for distributed group support. *Proceedings of the 32nd Hawaii International Conference on System Sciences*, Maui, HI, 5-8 January 1999. IEEE. doi: 10.1109/HICSS.1999.772718

Riley, J. M., Endsley, M. R., Bolstad, C. A., & Cuevas, H. M. (2006). Collaborative planning and situation awareness in Army command and control. *Ergonomics, 49 (12-13)*, 1139-1153. (Special Issue: Command and Control).

Salas, E., Cannon-Bowers, J. A., & Johnston, J. H. (1997). How can you turn a team of experts into an expert team? In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 359–370). Hillsdale, NJ: Erlbaum.

Stagl, K. C., Salas, E., Rosen, M. A., Priest, H. A., Burke, C. S., Goodwin, G. F., & Johnston J. H. (2007). Distributed team performance: A multi-level review of distribution, demography, and decision making. In F. Dansereau & F. J. Yammarino (Eds.), *Multi-level issues in organizations and time* (*Research in multi level issues*, Volume 6) (pp.11-58). Cambridge, MA: Emerald Group Publishing Limited.

Strater, L., Cuevas, H. M., Scielzo, S., Connors, E. S., Gonzalez, C., Ungvarsky, D. M., & Endsley, M. R. (2011). An investigation of technology-mediated ad hoc team operations: Consideration of components of team situation awareness. In K. L. Mosier & U. M. Fischer (Eds.), *Informed by knowledge: Expert performance in complex situations* (pp. 153-168). Boca Raton, FL: Routledge – Taylor & Francis.

Strater, L., McNeese, M., Cuevas, H. M., Fan, X., Oh, S., & Yen, J. Making sense of error patterns: Toward appropriate human-agent trust. Unpublished manuscript.

U. S. Army. (2001). *Concepts for the Objective Force: White Paper*. Author.

U. S. Department of Defense (1994). *DOD Joint Staff Publication No. 1-02* (JP 1-02): Department of Defense Dictionary of Military and Associated Terms. Author.

NEW AVIONICS TECHNOLOGIES HUMAN FACTORS

Michelle Yeh
Federal Aviation Administration
Washington, DC

Stephanie Chase
US DOT Volpe Center
Cambridge, MA

Juliana Goh
MITRE
McLean, VA

Bill Rogers
Honeywell
Golden Valley, MN

Nadine Sarter
University of Michigan
Ann Arbor, MI

The purpose of this symposium is to identify and address common human factors issues with new and emerging avionics technologies, share lessons learned, and to provide an understanding of how the Federal Aviation Administration (FAA) applies human factors research to enhance aviation safety. Flight deck technologies have been changing at a rapid pace, requiring updates to Federal Aviation Administration (FAA) regulations, guidance, and policy. This symposium will focus on flight deck technologies that will assist in NextGen implementation by improving flight crew awareness through Cockpit Displays of Traffic Information (CDTI), Airport Moving Maps, Primary Flight Displays (PFDs), and portable technologies. The results of this human factors research helps enable the FAA Office of Aviation Safety to develop and update establish evaluation criteria, operational procedures, and training recommendations.

Modern avionics are key to achieving NextGen's goals to increase efficiency, enhance safety, and improve situation awareness, both in the air and on the ground. New and emerging flight deck technologies have been changing at a rapid pace, and research is needed to identify the human factors/pilot interface issues associated with them. Pilot performance and efficiency during NextGen operations, support for the infrastructure needed to enable efficient and safe use of these advanced flight deck technologies, and the corresponding impact on operational safety are important considerations. The results of human factors research will help the FAA Aviation

Safety Organization establish evaluation criteria, operational procedures, and training recommendations.

The purpose of this symposium is to address new and emerging avionics technologies to highlight key research issues, share lessons learned, and discuss how the FAA uses the results of this research. This symposium will focus on flight deck technologies that will assist in NextGen implementation by improving flight crew awareness through Cockpit Displays of Traffic Information (CDTI), Airport Moving Maps, Primary Flight Displays (PFDs), and portable technologies. Common themes across multiple avionics projects include usability, system integration, symbology, controls, human/pilot error mitigation, workload, and distraction.

This symposium consists of five presentations. The first addresses the FAA's research needs associated with new avionics technologies, describe the general process for how research projects are initiated and coordinated, and how results are transitioned to implementation in the field or in regulation. The next four presentations address human factors challenges with new technologies focusing on symbology design, human interaction with information automation, information acquisition via electronic versus paper mediums, and the approval and evaluation of new technologies.

FAA Flight Deck Human Factors

The FAA recognizes human factors as a critical contributor for improving aviation safety and acknowledges that role in Order 9550.8, *Human Factors Policy* (FAA, 1993). This Order highlights that long-term improvements in aviation safety will be the result of consistent support for human factors research, analysis, and development and the implementation of those results (FAA, 1993). The role and implementation of human factors is spread throughout the Agency. The FAA Human Factors Division provides scientific and technical support for civil aviation human factors research. They manage, direct, and coordinate the FAA's human factors program for both flight deck and air traffic control. With respect to the flight deck, the FAA Human Factors Division provides research support to the FAA Office of Aviation Safety (AVS; Aircraft Certification and Flight Standards Services), who is responsible for the certification, production approval, and continued airworthiness of aircraft as well as the certification of pilots, mechanics, and others in safety-related positions. AVS performs two key activities: (1) evaluating and approving new or modified aircraft, equipment, operators, procedures, maintenance, etc. and (2) developing regulatory and guidance material. The goal of the FAA Flight Deck human factors research program is to provide research input to support these activities.

Human factors research into new technologies and operations enables a data-driven approach to the human factors aspects of new technologies and operations, analysis of safety data, and many other areas. This research is important for identifying potential or emerging safety issues, upcoming technologies and operations, and current operational safety issues, and to provide the research data to inform and support the AVS regulatory and oversight activities. Writing regulations and policy is difficult because it is important to write the guidance to say what is intended as well as to be clear as to what is *not* intended. In applying research products to the regulatory process, it is important for researchers to understand the role of AVS and what is needed, to use the same terminology or identify how it is different, and to provide data so AVS can be better positioned to know what should *not* be approved. The success of a research program is determined by how results of research are used. The benefit to the FAA is to develop and update establish evaluation criteria, operational procedures, and training recommendations. In some cases, industry may also benefit from the research results.

One example of a research success is the FAA's Electronic Flight Bag (EFB) research program. An EFB is an electronic display system that can be used to present data, such as electronic charts, checklists, documents, or to conduct basic calculations. EFBs can take many forms from a laptop or tablet computer to an installed display and processor in the flight deck. Figure 1 presents an overview of the steps from the initiation of research to implementation of a research program. The figure is intended to reflect the general process for initiating research and transitioning the results to implementation in the field or in regulation.

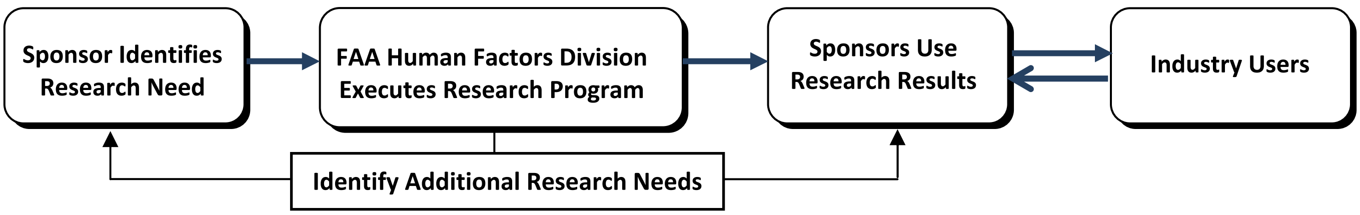


Figure 1. Research to Reality.

In the first block, technical sponsors in safety organizations within the FAA identify a research need and communicate that need to the research organization. For this particular example, the Electronic Flight Bag (EFB) research program was initiated by the FAA Human Factors Division in response to a research requirement from AVS to provide a capability for FAA Certification and Flight Standards personnel to evaluate human factors aspects of EFBs. Once this research requirement was identified, the FAA Human Factors Division sponsored the United States (US) Department of Transportation (DOT) Volpe Center to conduct this research, as noted in the figure by the second block. During the execution of this research program, Volpe Center researchers coordinated with FAA technical sponsors, the FAA Human Factors Division program manager, and the EFB industry, including EFB manufacturers, airline operators, and EFB software providers. The research products included a general reference on human factors considerations for EFBs, industry surveys to provide information on the state-of-the art, and EFB evaluation tool kits. (For a full list of EFB research results, see www.volpe.dot.gov/coi/hfrsa/work/aviation/efb.)

As a result of this coordination, AVS incorporated many of the results in policy and guidance material. This is Step 3 in Figure 1. The FAA referenced the Volpe Center EFB research in AC 120-76A, *Guidelines for the Certification, Airworthiness, and Operational Approval of Electronic Flight Bag Computing Devices*, and this information was also incorporated into subsequent revisions. The Volpe Center research was also used to develop the initial EFB Job Aid (Notice N8200.98). Notice N8200.98 has since expired, but the information was incorporated into FAA Order 8900.1, volume 4, chapter 15, Section 1, *Electronic Flight Bag Operational Authorization Process.* As a result, the Volpe Center research continues to influence any user who has sought approval of authorization to use an EFB because the research was referenced in FAA regulatory and guidance material.

In some cases, the research sponsored by the FAA may also support industry, as shown by the last block in Figure 1. As part of the Volpe coordination, industry participated in several efforts throughout the research projects to test Volpe-developed products. This collaboration benefited the FAA who received better research products as a result and industry who received human factors input on their products.

Examining the Intuitiveness of Traffic Symbology for CDTIs, Stephanie Chase, US DOT Volpe Center

The purpose of this research effort was to develop an understanding of which traffic symbol attributes are perceived to be useful by pilots and which symbol shapes and their properties (e.g., fill, shape, or color) are intuitive for pilots. The Volpe Center developed a paper-based questionnaire that comprised three tasks. In Task 1, pilots rated the usefulness of several attributes of traffic symbols. In Tasks 2 and 3, pilots were shown symbols from current and proposed symbol sets and asked to identify which symbols were representative of an information type or combination of information types. Seventy-nine pilots with a variety of flight experience participated in the study. The results indicated that color is an intuitive cue of airborne vs. ground traffic, although the shade of brown can be problematic since it is confusable with yellow at low brightness levels. Overall shapes indicating direction, such as a chevron, were found to be representative for directional information; non-directional symbols were found to be less intuitive and, in some cases, caused confusion. Other information types, such as data quality, generated mixed results. Initial findings indicate that alert information is perceived to be very useful to pilot while off scale and data quality less useful. Additional research is needed to further determine the usefulness of this information and how it relates to the intuitiveness for pilots.

Comparing Use of Electronic vs. Paper Documents on the Flight Deck, Juliana Goh, MITRE

EFBs provide a device to store and view documents, charts and maps that have traditionally been presented on paper. Given the increased use of EFBs on aircraft in commercial operations, it is important to understand any performance differences in how information is acquired, understood and retained when using electronic displays versus paper. It is also important to take into account different design related considerations when electronic displays are being used.

In general, acquiring textual information from electronic displays is at least as good as doing so from paper. Studies from the 1980s, based on extended reading from Cathode Ray Tube (CRT) displays showed, in general, that reading performance was slower than from paper. This difference could, however, be minimized and, in some cases, eliminated, when certain factors (e.g. double spacing, negative contrast) were taken into consideration. More recent studies based on better quality displays (e.g. Liquid Crystal Display) focused on variables (e.g. age and prior experience with electronic displays) other than display quality that could explain differences in reading performance from the two media. With regards to acquiring spatial information, the use of electronic maps have been demonstrated to be superior to paper maps because of the ability to selectively layer information when using an electronic display.

The research literature also suggests various design considerations to be made in the use of electronic displays: legibility, navigation and customization. Legibility refers to the clarity with which information is presented to the user. Navigation refers to the ease with which the user is able to locate him/herself within the body of text and go to locations within the text to retrieve the needed information. Finally, customization refers to the ability to interact with the electronic display or document in a manner that supports an individual's cognitive activities.

Information Automation, Bill Rogers, Honeywell

Flight Deck Information Automation (IA) as a distinct type of automation was proposed by Billings (1991). IA can integrate, summarize, distribute, format, abstract, prioritize, categorize, calculate, and process information in a variety of ways to support pilot tasks. It can include decision, task, and information management aids. With the proliferation of systems on the flight deck that are intended to provide information and situation awareness for the flight crew (e.g., EFBs, DataComm systems, advisory systems, decision aids, electronic charts, etc.), issues related to IA systems are likely to increase, but as a distinct type of automation, it is not well understood; often the term "automation" is used to describe automated systems of all kinds, including control automation, information automation, and management automation. Human Factors issues that are associated with control automation, such as mode confusions, may not be as important to IA, and certain types of human factors issues and pilot errors might be prevalent in interacting with information automation that are minimal for other types of automation.

FAA-sponsored Honeywell work focused on IA will be described here. The overall goal of this work is to provide recommendations for designers and evaluators of IA systems to assure that Human Factors issues unique to IA systems are identified and mitigated. Work will be described that defines IA and presents a framework for comparing and contrasting it to other types of automation. Specifically, a framework distinguishing different types of automation by utilizing human information processing stages, and a characterization of what entity is being controlled, will be described. Based on this automation framework, types of Human Factors issues for IA were identified and will be described, especially those that are hypothesized to be unique to IA or that likely manifest themselves in a substantially different way for IA than for other types of automation. Further, characteristics of IA that could impact user performance and user-IA system interaction, such as complexity and opacity, will be described in the context of potential risks and mitigations.

Finally, plans for empirical studies to evaluate the issues and mitigations identified analytically will be described, and examples of the intended outputs of the project will be provided.

Guidance and Tools for the Evaluation/Certification of NextGen Primary Flight Deck Displays, Nadine Sarter, University of Michigan

Aircraft technologies and operations can be expected to continue to grow in complexity. This trend brings with it an increase in the number of pilot responsibilities and the amount of data that is available to, and needs to be considered by flight crews. One important flight deck display that has changed significantly in recent years is the Primary Flight Display (PFD). More information has been added to this display (e.g., terrain information and synthetic vision), and a considerable number of different designs have been proposed. This project focuses on helping the FAA develop approval criteria for future PFDs. In particular, though not exclusively, the issue of clutter is being examined. Clutter has been defined as the result of high information density and/or poor layout of information. Performance effects of clutter are the result of an interaction between these display-related factors and top-down operator-related factors, such as experience. In an effort to provide guidance for the design and evaluation/certification of PFDs for advanced aircraft, we have conducted a survey of current PFD designs and another survey of pilots' operational experiences with these PFDs. We also reviewed and compiled existing research findings and regulatory documents that are relevant to the evaluation of PFDs. This has resulted in a draft general guidance documents and a PFD evaluation checklist for certification personnel. Finally, we are conducting simulation studies to develop eye tracking-based assessment tools that can detect the various attentional costs associated with clutter. These tools will be useful to manufacturers by informing the development and iterative refinement of PFDs as well as providing supportive evidence for the effectiveness of a proposed design.

References

Federal Aviation Administration. (1993). Order 9550.8, *Human Factors Policy*. Washington, DC: FAA.

Federal Aviation Administration. (2006). Notice N 8200.98, *Electronic Flight Bag Job* Aid. Washington, DC: FAA.

Federal Aviation Administration. (2012a). Advisory Circular (AC) 120-76B,*Guidelines for the Certification, Airworthiness, and Operational Approval of Electronic Flight Bag Computing Devices*. Issued on June 1, 2012. Washington, DC: FAA.

Federal Aviation Administration. (2012b). Order 8900.1, volume 4, chapter 15, Section 1, *Electronic Flight Bag Operational Authorization Process*. Washington, DC: FAA.

VISUALIZATION OF PAIRWISE CONFLICT RESOLUTION FOR AIR TRAFFIC CONTROL

M. M. (René) van Paassen
Aerospace Engineering – Delft University of Technology
Delft, The Netherlands

Air traffic capacity is mainly bound by Air Traffic Controller (ATCo) workload, which leads to problems in the view of the steadily increasing demand for air transport. Additional automation tools to support the ATCo in his current working practices are necessary. Visualization of control possibilities for aircraft by means of the "Solution Space" approach provides a first step in this direction. However, these visualizations focus on the control possibilities for a single aircraft, and a known problem is relating the indicated conflict back to the involved aircraft. This paper discusses the design of a visualization that shows the maneuvering options for a pair of aircraft in a conflict. As for the previous solution-space based display, the Ecological Interface Design framework is used to develop the design. The interface allows the ATCo to decide which aircraft should maneuver to most efficiently solve a conflict, or assists in selecting a joint maneuver, in which both aircraft make smaller maneuvers to solve the conflict. The manner in which the interface answers the requirements discovered with the work domain analysis and task analysis is discussed.

Currently, Air Traffic Controllers (ATCo's) perform a sector-based tactical form of control. They are responsible for planning and managing traffic within their assigned airspace, often with little help from automated tools. With the slow, but accumulating increase in air traffic, this makes the task of an ATCo a very demanding one, requiring an extensive selection process to find individuals capable of performing this task and a long training.

Limited automation is available, normally in the form of conflict detection probes or path prediction visualization. More recently, interfaces based on the visualization of "Velocity Obstacles" are being developed. Velocity Obstacles (VO) is a term originating from robotics – although similar theories considerably predate robotics (e.g. the Battenberg Course Indicator), and present the set of velocities of a robot (or, in our case a vehicle) that will result in a collision with another moving object. In aeronautics this has been labeled the "Solution Space", and interfaces employing this concept present the blocked velocities and heading for a selected aircraft (Mercado-Velasco, Mulder, & van Paassen, 2010; Abdul Rahman, van Paassen, & Mulder, 2011; Lodder, Comans, van Paassen, & Mulder, 2011).

Using such a representation, one can determine the safe heading and speed for the aircraft under control. However, in the case of conflicts involving two aircraft, which is by far the most common case, such a tool would be of immediate use only when the conflict is solved by maneuvering only one of the two aircraft. Judiciously applied one can also use the tool to solve only part of the problem with one aircraft, and then proceed by using the tool on the matching aircraft in the problem to complete the solution. This paper explores the possibilities to develop a visualization that that can support an ATCo in solving a two-aircraft conflict by having the two aircraft in the conflict both contribute with a maneuver.

Scope of the work

Air traffic control tasks differ considerably for different sectors. Large upper airspace sectors mainly deal with monitoring of overflying traffic. Most conflicts from crossing traffic are solved by assigning different altitudes to the traffic.

Approach sectors have less crossing traffic, and the main focus of the work is on departing and arriving traffic. Arriving traffic normally needs to be delivered to an arrival sector through one or more exit waypoints and most traffic needs to be brought to a single flight level (altitude) for the exit waypoint. Likewise, traffic that departs from an aircraft enters such a sector from one or a few entry points, and needs to be cleared for a climb to cruise altitude.

In the current practice, aircraft that are in a descent or climb need to be separated horizontally from all traffic at and between their current flight level and the flight level the aircraft has been cleared to climb or descend to. That is, the ATCo cannot make any assumptions about the climb or descent speed of an aircraft, and cannot assume that aircraft can cross

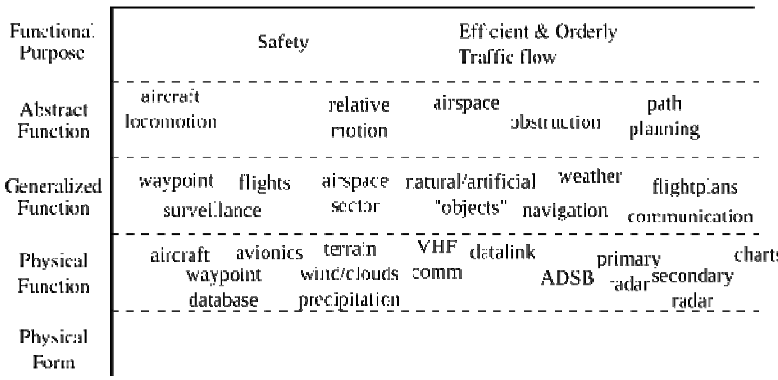


Figure 1: Work Domain Analysis for the task of Air Traffic Control.

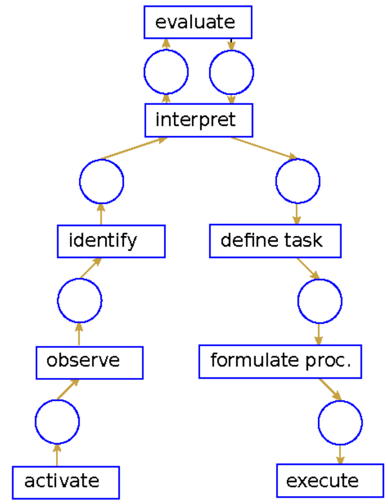


Figure 2: Rasmussen's decision ladder (Vicente, 1999)

with vertical separation, unless the ranges from current to assigned flight levels for both aircraft are disjunct with the required minimal separation between them. This makes horizontal separation a valid task for these situations, and one that cannot be always substituted by vertical separation solutions.

This paper focuses on the control task in horizontal separation. We consider control in the vertical dimension separate from this task, i.e., this control task may be applied to aircraft with overlapping vertical ranges in a descent or climb, or aircraft that is already brought to the desired altitude and it is not desirable or feasible to solve a conflict with vertical separation.

The main focus of the paper will be on the design of a support interface for horizontal separation of pairs of aircraft (which thus can include climbing and descending traffic with overlapping current and cleared flight level ranges), with the explicit possibility of using instructions to both aircraft to solve the conflict.

Work Domain Analysis

The results of the work domain analysis are given in an Abstraction Hierarchy. The AH summarizes knowledge on the work domain at different levels of abstraction. At the highest level, labeled "Functional Purpose" in this AH, the goals of the system are identified. The primary goal of air traffic control is to ensure the safety in the air. Only when safety is ensured, the ATCo can devote attention to the next goal, ensuring efficient and orderly flow of traffic (Figure 1). As with previous analyses for this domain, relative locomotion and absolute locomotion are salient functions at the Abstract Function level. A requirement for locomotion is the availability of airspace, and a function that impedes locomotion are dynamic and static obstructions, for example special use airspace, other aircraft or terrain. Although they could be considered as not belonging strictly to the work domain, but rather to be a tool involved in a specific solution, flight plans are included at this level too.

Control Task Analysis

The control task analysis for this system is performed for two different task, one is monitoring and maintaining the safety, corresponding to the first purpose in the AH, and the second is organizing the traffic flow. As a reminder, a skeleton for the decision ladder is given in Figure 2. The different interpretation of the actions in this ladder for the two tasks is given below. First for the safety task:

activate Starts with the recognition that a planned or current path of one or more aircraft will lead to a loss of separation in a short (5 to 10 min) time.

observe Information needs to be obtained about the distance/time to go until the conflict needs to be resolved, tracks of current aircraft in the vicinity, their plans.

identify Determine which aircraft are in conflict. Which aircraft affect or constrain the solution. How large is the conflict is, whether there is free space to solve the conflict, what will be the effect on the current flight plans.

interpret Determine what the disruption of the conflict is. How does it affect operation? What are remaining alternatives.

evaluate Choose the best or an acceptable option.

task definition Given the best possible solution, define what needs to be done to implement it. Which aircraft need to be maneuvered. What monitoring is necessary?

procedure formulation Determine the commands to give. Directions, sizes, new speeds/headings or altitudes.

execution Communicate with the pilots. Implement the solution. Monitor the follow-up.

For efficient flight execution, the stages in the decision ladder can be formulated as follows:

activate Starts with the recognition the planned or current path of one or more aircraft (or even the lack of having a planned path) will not bring the aircraft to its required exit point at the proper altitude.

observe Information needs to be obtained about the desired exit point, tracks of aircraft in the vicinity.

identify Determine the state of the current plan, find a possible path for bringing the aircraft to its exit point, determine whether crossing or competing aircraft form a limitation.

interpret Evaluate possible solutions and their effect on the traffic pattern and safety. Determine where in the sequence to place the current aircraft.

evaluate Choose the best or an acceptable option.

task definition Given the best possible solution, define what needs to be done to implement it. Which aircraft need to be maneuvered. What monitoring is necessary? Somehow store or record the plan.

procedure formulation Determine the commands to give. Directions, sizes, new speeds/headings or altitudes.

task execution Communicate with the pilots. Implement the solution. Monitor the follow-up.

Display Design

Previous conflict resolution displays for aircraft and for air traffic control were based on visualizing the relation between the relative velocity of an intruder (or conversely, the relative velocity towards an intruder) with the absolute velocity of the controlled aircraft itself. The current project has a different aim, in that we would like to use adjustment of the velocity of both aircraft to remove the conflict. The design needs to overcome a number of issues:

- **Control degrees of freedom.** Resolving a conflict in the horizontal plane with instructions to a single aircraft potentially requires two inputs; a new heading and a new speed. Thus, two degrees of freedom in the control vector. The current displays can show this information on a screen. Addition of a third degree, for example altitude control, has been attempted, but this requires the combination of several displays, which brings associated problems with maintaining visual momentum. Defining the four needed control inputs at the same time, i.e. heading change and speed change for the two aircraft, is not feasible. Somehow it should be possible to define a maneuver in which the two aircraft move in two steps.
- **Balanced maneuvering**. A way of reducing the number of degrees that have to be controlled for the two aircraft is by determining how much each of the aircraft contributes to the solution of the conflict. After choosing one of the vectors above, the ATCo has fixed how much of the solution must be provided by each aircraft.

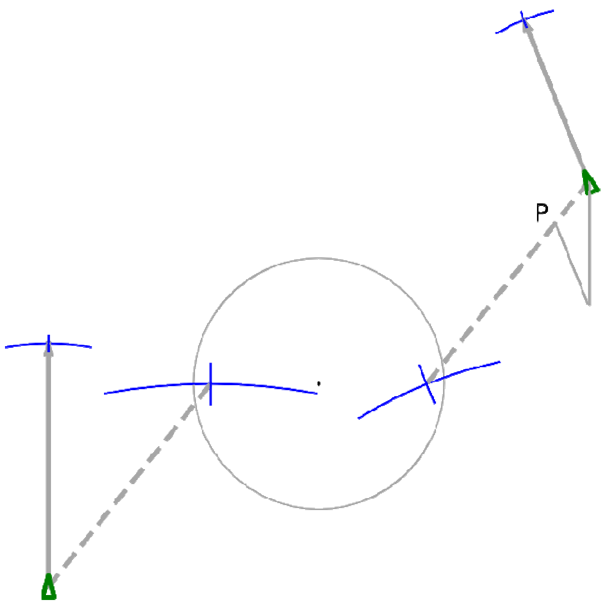


Figure 3: Diagram outlining the geometry in visualizing the conflict with respect to the midpoint between the two aircraft. Point "P", at one fifth of the 300 [s] velocity vector, indicates the relative velocity with a 60 [s] speed vector.

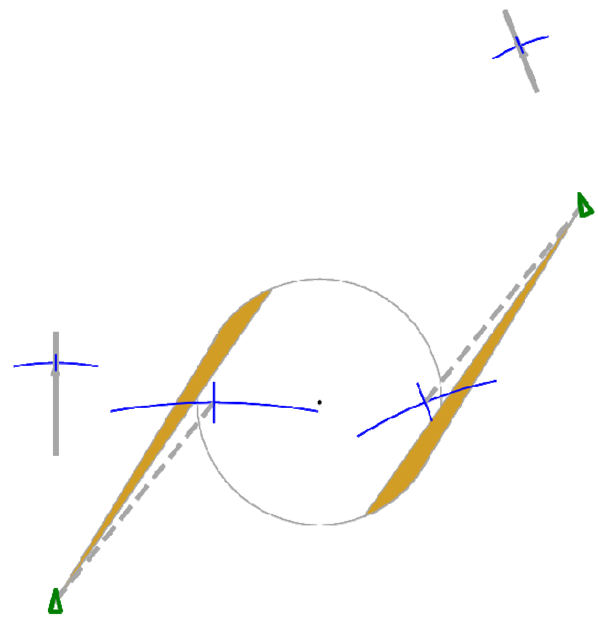


Figure 4: Further visualization of the conflict geometry. Triangular zones have been added, with one side indicating the change in relative velocity needed to solve half of the conflict, and the "outer" side indicating the relative velocity needed to solve the complete conflict.

- **Conflict priority and warning.** The SSD based displays discussed above provide a means to choose headings and speeds that resolve the present conflicts, but by itself the SSD does not aid in the detection of conflicts. To provide that functionality, the aircraft symbols in the plan view may be marked. However, with several marked symbols the task of prioritizing the conflicts still requires the ATCo to scan the different aircraft and call up their SSDs.
- **Traceability across abstraction levels**. The common format for an Air Traffic Display is a plan-view display with symbols representing the aircraft in the sector. Speed vectors or history dots ("breadcrumbs") can provide a velocity overlay. Any additional visualization is preferably traceable to the position and velocity physics of the aircraft, so that its role and effects become predictable and understandable. In other words, visualizations should be linkable to physical actions.

As a first step in the design, the aircraft-centric property of the SSD displays is questioned. The SSD displays provide a visualization of the relationship between the relative velocity and the absolute velocity of one considered aircraft of the aircraft in a conflict.

Instead, as shown in Figure 3, the center of a pair of aircraft that is in conflict (or indeed, any pair of aircraft) can be considered. When taking this midway between the two aircraft a point can considered that is always in the middle of the two aircraft, also when these are at their closest point of approach. The speed of this point is the average (vectorially) of the speeds of the two aircraft. Note that this has some relation to the visualization in (Gaukrodger et al., 2009), although there all conflicts are presented in the same space, making it difficult to relate conflicts to the aircraft involved.

Next, let's consider the relative velocity of the two aircraft with respect to this point. By subtracting the velocity of the midpoint from both aircraft velocities, the relative velocity is obtained. The two relative velocity vectors are parallel (the dotted lines in Figure 3). As was claimed before, the midpoint moves with the pair of aircraft. By considering the relative velocity with respect to the midpoint, one can see that relative position of the two aircraft at closest point of approach, is where a line at the midpoint, perpendicular to the line through the two current aircraft positions, crosses the relative velocity vectors. To have sufficient separation at that point, the distance between the vectors should be larger than the protected zone size, meaning that the relative velocity vectors should stay out of a circle with a diameter of the protected zone size centered on the midpoint, which we will term avoidance zone.

This already gives one option into visualizing the safety goal, see Figure 3. By indicating how much of the circle around the midpoint is clipped by the velocity vectors, or, alternatively, how much intrusion of this circle is present (Figure 4), the distance at closest point of approach is indicated. This visualization has the additional advantage of having a surface area roughly proportional to the severity of the conflict.

The visualization by itself covers detecting the problem, indicating which two aircraft are involved, and providing data for evaluation of safety (activate, observe, identify and part of the data needed for evaluate, Figure 2. However, if we want to use this visualization in an EID interface, we should somehow link the action possibilities, i.e. changing speed and heading of the individual aircraft, to the effects on the clipping of the midpoint circle.

In most cases, and when altitude changes are not applicable, a heading change will be preferred to a speed change. When considering a – possibly combined – heading and or speed change as a change in the aircraft's speed vector, the most efficient coordinated maneuver of the two aircraft, in velocity space, will be perpendicular to the relative velocity with respect to the midpoint. To support the ATCo in deciding on a maneuver, the "common" maneuvering options need to be elaborated somehow in the context of the constraints in the relative velocity space.

Given that the speed of the two aircraft can be represented by vectors v1 and v2 , the speed of the midpoint is simply $\underline{v}_m = (\underline{v}_1 + \underline{v}_2)/2$. The relative velocity of aircraft 1 with respect to the midpoint and its avoidance circle is $\underline{v}_{r1} = \frac{1}{2}(v_1 - v_2)$.

Similarly the relative velocity of aircraft 2 can be calculated. However, presenting speed vectors on a plan view display requires a scaling. A feasible "size" for the speed vector is 60 seconds, i.e., the speed of the aircraft is converted into distance by multiplying with 60 seconds and taking that for the speed vector size. For conflict detection and resolution, 60 seconds is too short, commonly 5 minutes is needed, so the relative velocity vector requires a scaling of 300 seconds. If the tip of the relative vector then points into or through the avoidance zone, this then indicates that there will be a loss of separation within 5 minutes.

On the other hand, a fixed scaling with 300 seconds might produced a confusing representation when the relative velocity of the two aircraft is very large, such as with a blunt angle crossing or a head-on conflict. In that case the closest point of approach may be much closer – in time – than five minutes. The visualization of the relative velocity then extends to a point way beyond the closest point of approach, possibly cluttering other parts of the display. The solution proposed for this is limiting the relative velocity vector size to the distance to the midpoint, symbols will have to indicate that the vectors are calculated for a shorter time than 5 minutes.

The scaling does complicate the link to the action possibilities. The easiest way to represent the action options for an aircraft is by expressing them as a change in the aircraft's speed vector. Half of that change carries over to the relative velocity vector (cf. the second equation), but the scaling of the relative velocity vector is – normally – larger than that of the velocity vector, so it carries over to a point closer by, point P in Figure 3.

The aircraft speed is also limited, and this limitation should be discoverable from the interface. To visualize that, the speed vector can be modified to show the range of achievable speeds, and the tip can be replaced by a cross, sized to represent 10 kts speed change (along the vector) and 10 degrees heading change (curved section perpendicular to the vector). To link this scaling to the tip of the relative speed vector, and thus to the separation in relative space, the same cross is repeated at that tip, but magnified to reflect the scaling of the relative velocity vector.

The visualization options for the display have not yet been finalized. Assuming the basic geometry, several options are still open and need to be tested in evaluations. The avoidance circle does not need to be visualized completely, only visualizing the parts of the circle that are cut out by the relative velocity vectors would suffice. This has the additional advantage that the size of the symbology will correspond to the urgency of the problem.

Comparison to Analysis

The avoidance zones in the display will pop up only when a conflict is detected, which means that with current heading and velocity the separation will be lost in 5 minutes. Such behavior adequately supports activation of the task of resolving conflicts and keeping separation. The degree of conflict, and the available means to solve it are visible in the depth of penetration of the conflict zone and the sizes and orientation of the action spaces. The location of the conflict zones also gives information on the aircraft involved in a conflict. Using the double rings, the maneuvering can either be distributed over both aircraft or assigned to a single aircraft. The visualization is most suited to conflict avoidance. Compared to the SSD displays, the visualization of the available action space is missing; with an SSD display, one can choose a "free" heading, with this display one can choose a heading that solves a particular conflict, but whether that heading introduces another conflict is not clear beforehand. This might lead to more exploratory – what if type – use of the interface. As an additional bonus, the gravity of a conflict roughly corresponds to the size of the visualization area, making pressing conflicts inherently more salient.

Conclusion and recommendations

Starting from the principle of visualizing relative velocity, a display presentation for – potentially cooperative – solving of pairwise aircraft conflicts is developed. The presentation supports the operator in several steps outlined in the cognitive task analysis; detection of a conflict (zones come up), and in the collection of information about time to go and involved aircraft. The visualization of "standard" actions helps in defining an avoidance strategy and determining which commands to give. Rather than focusing on a single aircraft to solve a conflict, the visualization focuses on the pair of aircraft in the conflict. Using a double boundary for the conflict solution, one halfway to full separation, and the second indicating full separation, the operator will have a choice to solve a conflict with one or with two aircraft.

In the current design, the functional purpose level, for example by showing where the aircraft should be guided, is not explicitly shown. This information is normally shown as extra information in the aircraft label or in the flight plan, i.e. as symbolic information. Different visualization options can still be explored, and the display needs to be evaluated in simulation. In addition, the vertical dimension needs to be added; even when supporting the current practice in ATC, which is to treat climbing or descending traffic as if it were occupying multiple flight levels, climbs or descents will have a significant influence on the aircraft's speed range.

Acknowledgements

Ideas for this visualization were partially formed in discussions on related display design projects by MSc and PhD students, for which I particularly want to thank Joost, Rolf, Max, Clark and Jan. The interface was prototyped in a 250-line python program using numpy and matplotlib.

References

Abdul Rahman, S. M. B., van Paassen, M. M., & Mulder, M. (2011, August). Using the solution space diagram in measuring the effect of sector complexity during merging scenarios. In Aiaa guidance, navigation and control conference (pp. 1–25). Portland (OR). (AIAA-2011-6693)

Gaukrodger, S., Wong, W., Fields, B., MLoomes, M., Han, F., Mun˜ oz Araco´ n, M., . . . Monteleone, A. (2009, December). The multi-conflict display. In 8th innovative research workshop & exhibition (pp. 93–100). Bretigny, France: EUROCONTROL.

Lodder, J., Comans, J., van Paassen, M. M., & Mulder, M. (2011, April). Altitude-extended solution space diagram for air traffic controllers. In J. Flach, M. Vidulich, & P. Tsang (Eds.), Proceedings of the 16th international symposium on aviation psychology (p. 345-350). Dayton (OH).

Mercado-Velasco, G., Mulder, M., & van Paassen, M. M. (2010, August). Analysis of air traffic controller workload reduction based on the solution space for the merging task. In Aiaa guidance, navigation and control conference (p. 18). Toronto, CA. (AIAA-2010-7541)

Vicente, K. J. (1999). Cognitive work analysis. toward a safe, productive, and healthy computer-based work.Mahwah, NJ: Lawrence Erlbaum Publishers.

EVALUATING ASAP (ANTICIPATION SUPPORT FOR AERONAUTICAL PLANNING): A USER-CENTERED CASE STUDY

Sami Lini[1,2,4]
Christophe Bey[1,2,4]
Sylvain Hourlier[4]
Bruno Vallespir[1,3]
Axel Johnston[1,2,4]
Pierre-Alexandre Favier[2,4,5]

[1] Univ. Bordeaux, IMS, UMR 5218, F-33400 Talence, France
[2] CNRS, IMS, UMR 5218, CIH/ISCC team, F-33000 Bordeaux, France
[3] CNRS, IMS, UMR 5218, LAPS Department, F-33400 Talence, France
[4] HEAL (Thales Avionics/ENSC), F-33000 Bordeaux, France
[5] Univ. Bordeaux, IPB, ENSC, F-33000 Bordeaux, France

During ISAP 2011, we presented an early work about the theoretical background for the design of an anticipation support for civilian pilots. Two years later, the ASAP (Anticipation Support for Aeronautical Planning) project has grown and a preliminary version of our anticipation support has been implemented. Following a cognitive engineering approach, users were involved at each step of the design. An intermediate study allowed refining the information processing in order to better fit the actual need. On the basis of interviews and activity analysis, functionalities were implemented aiming at improving anticipation skills. In the context of a flight simulator and along two scenarios using the ASAP interface pilots are asked to act the way they are used to. In this paper evaluated functionalities as well as experimental protocol are detailed. Lessons learned are presented and discussed. Conclusions are drawn for future developments.

During ISAP 2011, we presented an early work about the theoretical background for the design of an anticipation support for civilian pilots. Two years later, the ASAP (Anticipation Support for Aeronautical Planning) project has grown and a preliminary version of our anticipation support has been implemented.

Following a cognitive engineering approach, users were involved at each step of the design. An intermediate study allowed refining the information processing in order to better fit the actual need. On the basis of interviews and activity analysis, functionalities were implemented aiming at improving anticipation skills. Evaluating such an interface only made sense with actual users' expertise.

In the context of a flight simulator and along two scenarios using the ASAP interface pilots are asked to act the way they are used to. A cognitive and a functional evaluation are conducted at the same time. Hypotheses are the following: the ASAP HMI allows an improvement of situational awareness; it allows a decrease in cognitive workload as well as an improvement of overall performance.

Situation awareness is assessed using 3D-SART. Workload is assessed using physiological data (pupil diameter), real-time probe and subjective measures (NASA-TLX). A performance index is also analyzed.

In this paper evaluated functionalities as well as experimental protocol are detailed. Conclusions are drawn for future developments.

Introduction

Anticipation

Anticipation is often considered as level 3 of Endsley's situation awareness model (Endsley, 1995): the one in which environment information is projected in time in order to assess future states. Cellier (1996) gives the following definition about anticipation: an "*activity consisting of evaluating the future state of a dynamic process, determining the type and timing of actions to undertake on the basis of a representation of the process in the future, and, finally, mentally evaluating the possibilities of these actions. It is dependent on the overall goal assigned to an operator in a dynamic situation, which is to keep the process, physical or otherwise, within acceptable limits, and therefore avoid the propagation of disturbances. It is also governed by a logic aimed at reducing the complexity of a situation. Finally, it is a way of managing individual resources"*. This definition is close to the acceptation of an anticipatory system in computer sciences, which Rosen (1985) defines as following: "*a system containing a predictive model of itself and/or of its environment, which allows it to state at an instant in accord with the model's predictions pertaining to a later instant"*.

Hoc (1987) highlights the close link between abstraction ability and expertise: what distinguishes experts and naïve subjects is the capacity to abstract data from a problem and fit them into a generic frame it will specify. This result gives credit to the hypothesis according which anticipation is an ability that defines experts. In our context, it justifies the approach aiming at providing pilots, particularly less experienced ones, a means to compensate a potential lack of anticipation ability.

In the aeronautics field, pilots are taught to be "*in front of their plane*" and are encouraged to "*make permanent assumptions about the future situation in order to actively adapt to this situation and not wait for it to occur*" (Amalberti, 1996). Anticipation is a major skill that pilots are taught to develop all along their career.

Design process

Since the very beginning, we put the stress on putting actual end users (i.e. pilots) in the design loop. A first video recording was made in the cockpit during the descent and approach phases of an actual flight. The recording was analyzed and an activity analysis was performed. A semi-direct interview of this very pilot was conducted to validate our activity model. A second pilot was then involved to double check this modeling. This led to identifying anticipation material and the critical points linked to it.

During the process, two major approaches were identified to help better anticipate:

- managing tasks which need to be performed during the flight,
- considering a change of trajectory, during a divert for instance.

These two trends were presented to a set of pilots who approved them. It was then decided to implement:

- a task panel presenting an improved check/do list,
- a vertical display with augmented information about trajectory,

- a map display with augmented information about trajectory and divert planning options.

A hierarchical task analysis allowed defining the tasks graph and an algorithm was implemented to compute in real-time the feasibility margins for every single task with observation of the whole list of flight constraints (see Lini, Vallespir, Hourlier, Labat, & Favier, 2013). This algorithm allows the computation of a whole new trajectory in case of a diversion and the presentation of the matching check/do list with its feasibility margins.

Because of a non-disclosure agreement with Thales Avionics due to ongoing patents, it is not possible to show the ASAP HMI yet.

Objectives and hypotheses

As anticipation is assumed to be a major part of situation awareness and mental workload management, the objective of this work is to assess the impact of our anticipation support on both the situation awareness and the cognitive load of actual pilots. Two groups were compared: in the first one, our interface was available during the first scenario and not during the second. The second group was organized the other way around.

Experiment plan

As part of the ASAP (Anticipation Support for Aeronautical Planning) project, we chose to focus our experiment on the descent and approach phases when landing a civil aircraft. For this study, we chose to study a flight from Paris landing at Rio de Janeiro airport. The final approach is known to need anticipation, due to landscape.

Task description

The scenarios were played out in a civil aviation simulator (Figure 1). Pilots were asked to interact with the cockpit the way they are used to in their day-to-day cockpit environment. They were asked to follow the descent trajectory shown on the navigation display using only the automatic pilot (heading and vertical speed vector). During the final approach, they were allowed to use the ILS' localizer and glide.

This is not a very usual activity for them as they are more used to fly on a FMS driven basis and mainly monitor what is happening. We made the choice to ask them to "manually" follow the track in order to keep them in the perception/action loop.

Figure 1: Overview of the simulator: the pilot flying (participant) was seating on the right, while a pilot non-flying was seating next to him on the left. The experimenter was monitoring from the left side of the room. Cockpit information was shown on the front panel while ASAP HMI was shown on the central upper screen and the 3D-SART on the central downer screen.

The two scenarios were designed as follows:

- Scenario 1 was a nominal descent on Rio de Janeiro International airport, coming from Paris. The scenario started at the top of descent and ended when the plane had stopped on runway 10. Information about the aircraft position and the remaining petrol was given. Remaining petrol was calculated so as in case of a go-around it would only be possible to wait for 10 minutes.

- Scenario 2 presented the same initial conditions. The scenario started at the top of descent. 5 minutes after starting the descent, subjects got a message from the approach control telling them that runway 15 was closed and runway 10 was currently engaged. They were asked to prepare for a holding pattern. 5 minutes later, the information was confirmed and they were asked to wait for 10 minutes. The scenario ended when the pilot decided to divert.

Material and task description

The ASAP interface and the 3D-SART questionnaire were implemented in JAVA. The simulator used was based on the behaviour of X-Plane 9. All of the graphics for the simulator were generated using 8 PCs. The interface between the various software components on each of the PCs and X-Plane was implemented using an API developed in-house by Thales Avionics.

The ASAP interface and the 3D-SART questionnaire were projected on two 15.4-inch touch-screen placed to the subject's left, within reach. The adjustment buttons were provided on two modules: Saïtek Pro Flight Switch Panel and Saïtek Pro Flight Multi Panel, both located within reach to the subject's right. The outdoor view scrolls on a 3-m-diagonal screen placed 1 m from the subject. The cockpit was presented in a 15.4-inch screen.

Both this screen and ASAP's were captured in real-time using two synchronized AVerMedia HD cards. A GoPro Hero 2 was used as a scene camera and recorded at the same time on a third AVerMedia capture card.

36 pilots (Air France and business jet pilots) were recruited, 2 women and 34 men. The results for 3 subjects were excluded from analysis due to technical problems encountered during test runs. The first group was composed of 17 commercial pilots, aged between 28 and 52, accumulating between 3,000 and 16,000 flight hours (M = 7,670, σ = 3,900). The second group consisted of 16 commercial pilots, aged between 26 and 53, accumulating between 2,000 and 17,000 flight hours (M = 7,610, σ = 4,100).

Running test

After introduction of the system, subjects were trained for its use and allowed to familiarise themselves with it.

Situation awareness was measured using the 3D-SART questionnaire: throughout each scenario, the subject was asked to regularly answer a 3D-SART questionnaire. Every two minutes, the questionnaire was shown and a concomitant sound alert –phone ring- rang to warn the pilot.

Cognitive load was measured using three independent, crossed methods (Cegarra & Chevalier, 2008):

- Pupillometry: it is an indirect indicator of cognitive load (Beatty & Lucero-Wagoner, 2000). It is measured by continuously monitoring pupil dilation. To do so, subjects were fitted with a mobile eye tracker, Tobii's glasses, a head-mounted eye tracking system resembling a pair of glasses. The tracker is monocular (right eye only), sampling at 30 Hz with 56° × 40° recording visual angle. Tobii studio, the data processing software, allows dealing with mean pupil dilation, i.e. the percentage of dilation compared to the mean dilation measured during the calibration phase.
- Subjective measurements: at the end of each test mode, the subject was asked to subjectively assess his/her cognitive load using the French version of the NASA-TLX questionnaire (Cegarra, 2009).
- Real-time probe: the subject was asked to answer the 3D-SART as quickly as possible.

Scenario 1 and scenario 2 were played one after the other in that order. During scenario 2, the time taken to make a decision and divert will also be analysed. STARs and approach charts were given prior to the flights and pilots took time to prepare them. They were then asked to perform a descent briefing the way they are used to. As performance indexes, it was checked during the flight if constraints from the charts are satisfied. Time taken to divert during scenario 2 is also analysed.

All these results feed our dependent variables.

Lessons learnt and first conclusions

This five-month long simulation campaign led to several lessons learnt about our setup, the protocol and a few other things:

1. However unrepresentative of an actual airplane our simulator was, it has on the contrary proved to be an asset for our study: the pilots who came were flying Airbus' planes or Boeing's (and a few others). All of them had to deal with the system the same way.

2. The use of a head-mounted eye-tracker proved to be much more complicated than expected. People suffering from presbyopia and used to wearing glasses are troublesome from this viewpoint: they are used to correct their vision on the short range and not on the long range. Thus, they are used to looking out of their glasses half of the time and therefor they tend to do so with the head-mounted eye-tracker.

3. The third conclusion is about the interest shown by the pilots' community towards both the study and the ASAP HMI: 36 commercial pilots came, sometimes from hundred kilometers to spend half-a-day on a voluntary basis just to share their expertise in order to improve their work environment.

4. The use of 3D-SART regularly shown throughout scenarii can be used to draw global trends about how pilots evaluate the situation complexity along time.

These lessons learnt will be used for further investigation about how to improve pilots' world of work.

References

Amalberti, R. (1996). *La conduite de systèmes à risques*. Paris: Presses Universitaires de France.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. *Handbook of psychophysiology* (2nd ed, pp. 142–162). New York, NY, US: Cambridge University Press.

Cegarra, J. (2009). Étude des propriétés de la version francophone de NASA-TLX. *EPIQUE*.

Cegarra, J., & Chevalier, A. (2008). The use of Tholos software for combining measures of mental workload: Toward theoretical and methodological improvements. *Behavior Research Methods*, *40*(4), 988–1000. doi:10.3758/brm.40.4.988

Cellier, J. M. (1996). Exigences et gestion temporelle dans les environnements dynamiques. *La gestion du temps dans les environnements dynamiques* (pp. 20–48). Paris: Presses Universitaires de France.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64.

Hoc, J. M. (1987). *Psychologie cognitive de la planification*. Presses Universitaires de Grenoble.

Lini, S., Vallespir, B., Hourlier, S., Labat, F., & Favier, P.-A. (2013). A cognitive engineering approach for showing feasibility margins on an in-flight planning. *Proceedings of the 17th International Symposium on Aviation Psychology*. Dayton, OH.

Rosen, R. (1985). *Anticipatory systems: philosophical, mathematical, and methodological foundations*. Pergamon.

UAS in the NAS: Survey Responses by ATC, Manned Aircraft Pilots, and UAS Pilots

James R. Comstock, Jr.
NASA Langley Research Center

Raymon McAdaragh
SGT, Inc., Hampton, VA

Rania W. Ghatas
NASA Langley Research Center

Daniel W. Burdette
Northrop Grumman Technical Services

Anna C. Trujillo
NASA Langley Research Center
Hampton, VA

NASA currently is working with industry and the Federal Aviation Administration (FAA) to establish future requirements for Unmanned Aircraft Systems (UAS) flying in the National Airspace System (NAS). To work these issues NASA has established a multi-center "UAS Integration in the NAS" project. In order to establish Ground Control Station requirements for UAS, the perspective of each of the major players in NAS operations was desired. Three on-line surveys were administered that focused on Air Traffic Controllers (ATC), pilots of manned aircraft, and pilots of UAS. Follow-up telephone interviews were conducted with some survey respondents. The survey questions addressed UAS control, navigation, and communications from the perspective of small and large unmanned aircraft. Questions also addressed issues of UAS equipage, especially with regard to sense and avoid capabilities. From the ATC and military ATC perspective, of particular interest is how mixed-operations (manned / UAS) have worked in the past and the role of aircraft equipage. Knowledge gained from this information is expected to assist the NASA UAS in the NAS project in directing research foci thus assisting the FAA in the development of rules, regulations, and policies related to UAS in the NAS.

The NASA "UAS Integration in the NAS" project is tasked with facilitating the process of developing the rules, regulations, and requirements needed to safely fly UAS of a variety of sizes and capabilities in the NAS. The U.S. General Accountability Office (2012) recently published a status report of progress towards integration efforts led by the FAA towards UAS Integration. A UAS Access Research and Development roadmap has also been developed by the NASA Langley Research Center (Verstynen, Foggia, & Hoffler, 2010). Key to the success of having UAS fly in the NAS, regardless of their size, is attention to the human factors issues of the Ground Control Station (GCS). The U. S. Department of Defense (2012) has published a GCS Human-Machine Interface Development and Standardization Guide, and other publications (e.g., McCarley & Wickens, 2005) have focused on the human factors issues of UAS in the NAS.

The purpose of the present paper is to present preliminary findings from on-line surveys that were conducted sampling the three major players involved when UAS are flying in the NAS. The surveys were targeted at ATC, including military ATC, pilots of manned aircraft, and UAS pilots. The surveys assessed the participant's background and experience in their particular area, followed by questions asked of all three groups as well as questions unique to the ATC, manned aircraft pilot, and UAS pilot operational domains.

Methodology

Separate on-line surveys were created and administered to Air Traffic Controllers (ATC), pilots of manned aircraft, and pilots of UAS. These on-line surveys were hosted on web-based SurveyMonkey. Survey content was reviewed by the NASA Langley Research Center Institutional Review Board and a "*Human Subject Research Volunteer Informed Consent Statement*" was presented at the beginning of the survey, with the survey respondent having to select "I Agree" in order to continue on to the survey questions.

For respondents there was a two step process, first signing on to the NASA Langley Human Subjects Recruitment website, and registering with contact information. Second, they would then receive an Access Code which would need to be entered on the survey website. This Access Code permitted a NASA Langley human subjects recruiter to pay subjects who were eligible to be paid (non-government, non-military), and to provide contact information for follow-up interviews, while keeping all other identifying information out of the response data files. To recruit survey respondents, the human subjects recruiter sent targeted emails to organizations identified by the research team (e.g., FAA, selected military bases, general aviation and commercial pilots, and selected manufacturers). There was a set of questions that were asked of all three groups as well as questions unique to each of the groups. The survey for ATC had 48 questions, the survey for pilots of manned aircraft had 46 questions, and the survey for UAS pilots had 72 questions. There were text box comment fields for most questions. Most respondents (90.5%) indicated that they would be willing to participate in a follow-up telephone interview.

For the ATC group, usable responses were obtained from 8 persons (5 male, 3 female), with a range of years as a Certified Professional Controller from 0 – 36, with a median of 7 years. For the manned aircraft pilot group usable responses were obtained from 27 persons (26 male, 1 female). The UAS pilots group was comprised of 9 persons (8 male, 1 female). Because of the limited space available for this paper, only selected results will be presented here.

Questions asked of all groups

Figure 1 shows the responses from each of the three groups to the question "Should the rules and requirements for the various classes of controlled airspace (Classes A, B, C, D, E, & G) be the same for UAS operations as they are for manned aircraft?" An interesting finding here is the dropping in "yes" responses from both Manned Aircraft Pilots and ATC towards the Class E and G Airspace, while the UAS Pilots did not show this change. This is also interesting in that many of the UAS pilots who responded have also been Manned Aircraft pilots. This may reflect that UAS pilot group expects to meet whatever rules and requirements there are for a given Airspace.

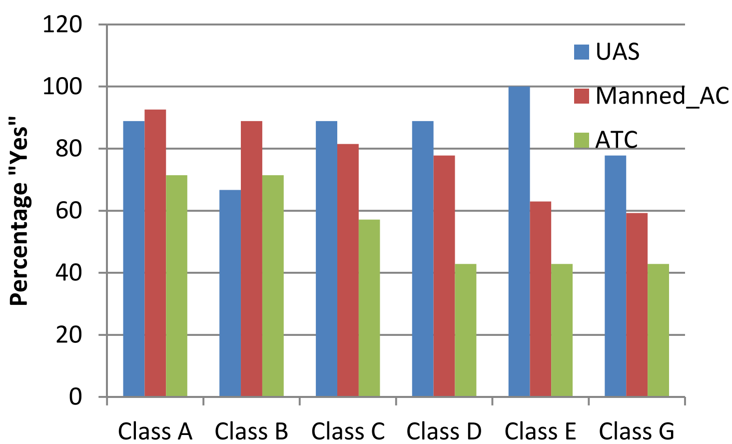


Figure 1. "Should the rules and requirements for the various classes of controlled airspace (A, B, C, D, E, & G) be the same for UAS operations as they are for manned aircraft?"

Figures 2, 3, and 4, show responses to related questions concerning the need for separate or special airspace, depending on the size and equipage of the aircraft. In Figure 2, the aircraft description is a small UAS without ATC communications and not transmitting position information. In this case the figure shows that some 58% of ATC respondents, about 80% of manned aircraft respondents, and about 50% of UAS pilots indicated "agree" or "strongly agree" to this statement. However, there were some UAS pilots who "strongly disagree."

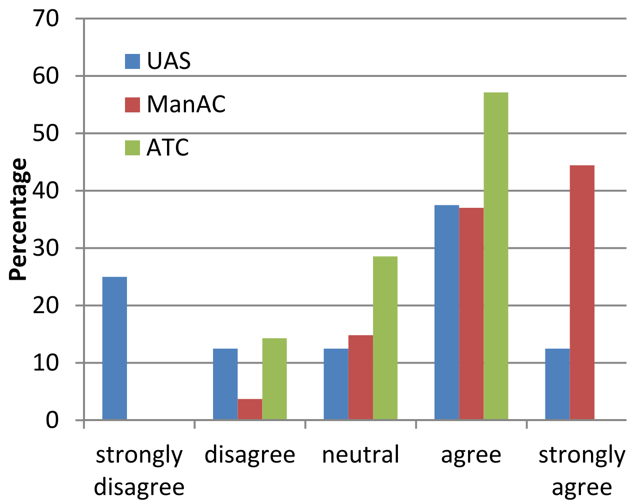


Figure 2. "I believe that small UAS (under 55 lbs) without ATC communications and without transmitting position (ADS-B) information will need separate or special airspace for their operations."

Figure 3 asks the same question but for small UAS "with ATC and transmitting position information," and the responses shift dramatically towards "disagree" or "strongly disagree" with regard to needing separate or special airspace. This result shows the importance of equipment that provides information that will allow the UAS to be "seen" and "communicated with" on perceptions of whether separate or special airspace will be needed.

Figure 4 shows the responses for medium and large UAS (> 55 lbs) with communications and transmitting position information. While there was 100% "disagree" or "strongly disagree" for the UAS pilots, and just over 50% disagreement for manned aircraft pilots, the ATC respondents were nearly evenly divided on the agree / disagree continuum. It is interesting that for the ATC group, the "agree" category was much higher for the medium and large UAS than for the similarly equipped small UAS shown in Figure 3. This may reflect a weighting of operational differences (e.g., Airspace Classes, airports needed) between the small and larger UAS in the response to this question for the ATC group.

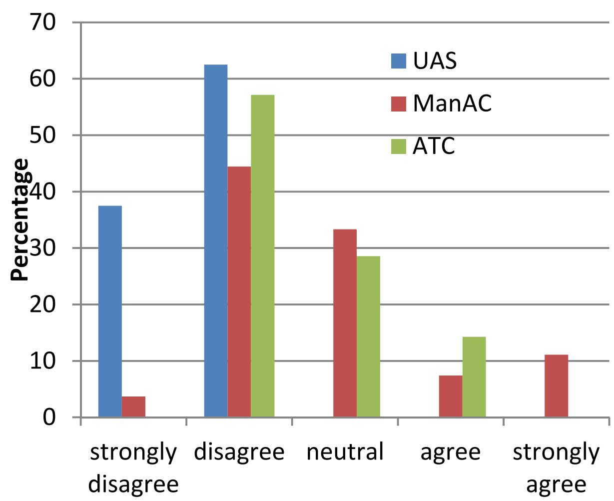


Figure 3. "I believe that small UAS (under 55 lbs) with ATC communications and transmitting position (ADS-B) information will need separate or special airspace for their operations."

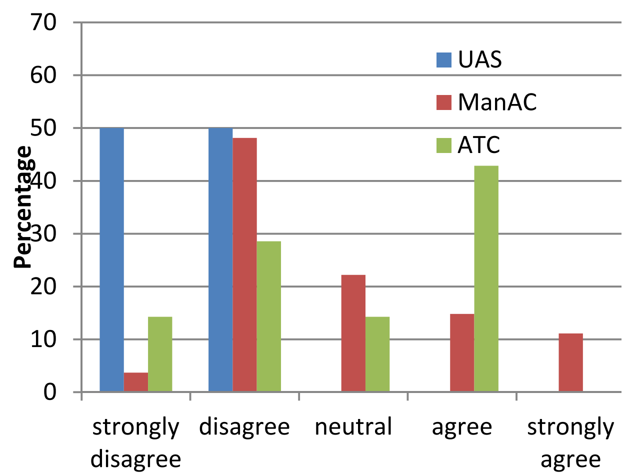


Figure 4. "I believe that Medium and Large UAS operating in the NAS with ATC communications and transmitting position (ADS-B) information will need separate or special airspace for their operations."

Manned Aircraft Pilots Questions

Several questions on the survey for pilots of manned aircraft addressed display of these aircraft on traffic displays and overhearing communications between ATC and these aircraft, also known as "party-line" information. Figure 5 shows the results for the question "When flying in an area in which UAS Operations are being conducted, how important is it to know that an aircraft shown on a Cockpit Display of Traffic Information (CDTI) is unmanned? (e.g., through symbology or data-tag information).” As shown in Figure 5, 20 of the 27 respondents (74%) rated this information as either “desirable” or “essential.” Two pilots, commenting on this question, said they needed to know if the UAS has TCAS (Traffic alert and Collision Avoidance System) and will automatically respond to an RA (Resolution Advisory). Another comment said that knowing the traffic aircraft was unmanned was more important if it was not able to respond to TCAS.

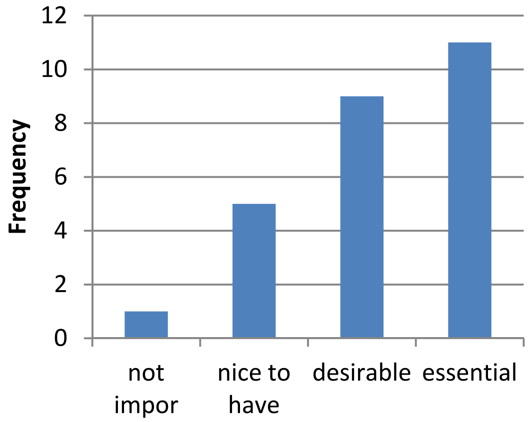


Figure 5. “When flying in an area in which UAS Operations are being conducted, how important is it to know that an aircraft shown on a Cockpit Display of Traffic Information (CDTI) is unmanned? (e.g., through symbology or data-tag information)” – manned aircraft question

In Figure 6 the results are shown for the question “When flying in an area in which UAS Operations are being conducted, how important is it that you hear ATC communications with the unmanned aircraft pilot? (sometimes referred to as the “party line”)” The responses here show 23 of 27 respondents (85%) indicated having “party line” information was “desirable” or “essential” and no one indicated that it was “not important.” Comments to this question said having this information: (1) was part of total situation awareness; (2) is a way to know if the UAS is responding appropriately to ATC and operator input; and, (3) is another trap for errors such as a clearance given in error or misunderstood that another set of ears might act as a barrier against.

Figure 7 presents the results for the question “If you are flying 1000-3000 ft Above Ground Level (AGL) in an area in which small UAS (under 55 lbs) are operating below 400 ft AGL, how important is the display of that aircraft on a Cockpit Display of Traffic Information (CDTI) display?” The responses here show 18 of 27 (66%) indicating that this information would be “desirable” or “essential”, while 4 of 27 (14.8%) indicating that this information was “not important.” Comments to this question noted that: (1) this information would be vital for altitude separations less than 1000 ft; (2) small UAS (under 55 lbs) would be nearly impossible to see air-to-air; and, (3) this

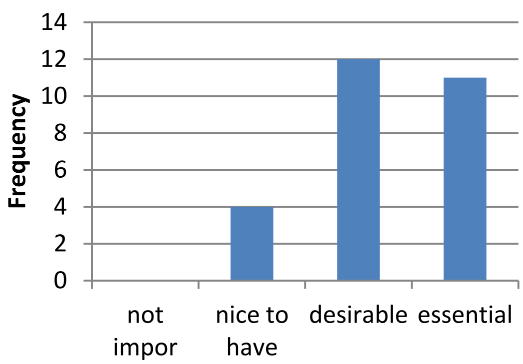


Figure 6. “When flying in an area in which UAS Operations are being conducted, how important is it that you hear ATC communications with the unmanned aircraft pilot? (sometimes referred to as the “party line”)” – manned aircraft question

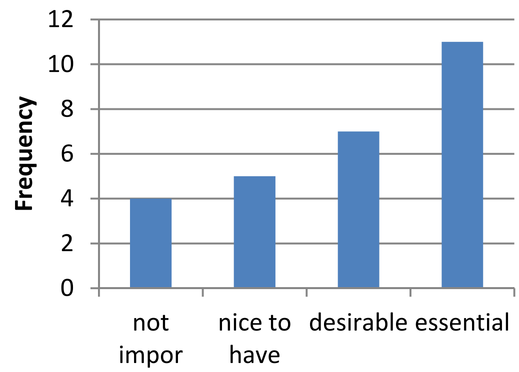


Figure 7. “If you are flying 1000-3000 ft Above Ground Level (AGL) in an area in which small UAS (under 55 lbs) are operating below 400 ft AGL, how important is the display of that aircraft on a Cockpit Display of Traffic Information (CDTI) display?” – manned aircraft question

information, while valuable, may not be available on many older generation General Aviation aircraft.

Selected ATC Questions

The following question was asked on the ATC survey: "When working manned aircraft and UAS in your airspace of responsibility, how important would it be to know that an aircraft shown on your radar display is unmanned? (e.g., through symbology or data-tag information)." ATC respondents were evenly divided between "desirable" and "essential," with no responses of “not important” or “nice to have.” Comments said this information would be “essential” for many reasons, including the lack of maneuverability and climb rate of UAS aircraft, as well as their inability to see and avoid, thus making this information a controller will need to make decisions regarding traffic calls, separation, and sequencing.

A related question concerning display of small UAS was "When you are working aircraft in your airspace of responsibility, in which small UAS (under 55 lbs) are operating below 400 ft AGL and more than 3 miles from an airfield, how important is the display of that aircraft (data tag information) on your radar display?" No ATC respondents indicated that this was "essential" information. The response of highest frequency was "desirable" (57%), followed by "not important" (28%) and "nice to have" (14%). Comments included: (1) Aircraft in Class D airspace can have arrival/departure route/pattern altitudes as low as 500 ft AGL; (2) in general, not important unless those operations are conducted within Class D airspace or in close proximity to the traffic pattern of any airports; and, (3) below 400 ft makes it generally safe, but would still want to know they were there to give traffic to low-operating aircraft/helicopters.

Selected UAS Pilot Questions

When viewing these results, keep in mind that the respondents indicated that they have experience with UAS of differing sizes and equipages. There were a number of open-ended questions addressing the GCS and related issues. One of these was “What sensory cue information, not provided currently, would help improve your situation awareness of the environment of the aircraft, the integrity of the aircraft’s flight and its mission?” Comments included: (1) “integrated displays for traffic (from ATC), TCAS, GCAS (ground collision avoidance system), weather;” (2) “Pilot's view camera;” (3) “weather radar;” (4) “Being able to see other traffic and weather surrounding UAS;” (5) “Audible cues would be helpful if the UA is not instrumented adequately;” and, (6) “Spoken messages.”

The next question was “How often is the UAS camera system used for navigation purposes?” Responses included (1) never (most frequent response), (2) "almost never, it can be useful in the terminal area," (3) "For emergencies only," (4) "Often, especially to avoid weather," and (5) "Whenever any clouds or precipitation is proximate" A related question was “How often is the UAS camera system used for “see and avoid” purposes?” The responses here included (1) Never or rarely, (2) “Never, the field of view is too wide,” (3) “Taxi only, and it only views forward,” (4) “Weather avoidance primarily, not traffic,” and, (5) “almost always during takeoff, departure and then during approach and landing.”

In answer to the question “Can a single UAS pilot perform all the tasks necessary to fly safely in the NAS?” the responses were 7 “yes” to 1 “no.” Responses to the question “How frequently during a typical mission are you in contact with ATC or other aircraft?” yielded 8 of 9 respondents indicating either “occasionally” or “routinely.” In response to the question “How frequently does the UAS automation do something unexpected?” answers were evenly distributed across “never,” “rarely,” and “occasionally,” but no one indicated “routinely.”

The following question addressed voice communications and communications latency: “If there is voice communications in the GCS, what could be improved to enable better voice communications, and has latency or delay in voice communications been a problem?” Responses included: (1) better radio equipment, (2) a second or third radio instead of just one, (3) second radio and radio selector, (4) no problems with latency or delay, (5) a faster link with a higher bandwidth; latency and delays are always a problem; faster link decreases the amount of processing the aircraft does with the voice signal, (6) latency is not normally a problem, however sometimes signal quality can be poor, and (7) latency is only a problem when the radios are busy and operations are by satellites; it can be hard to break in to make a call.

Implications for UAS design

Based on the survey responses and information from the follow-up interviews, there are two areas that will be briefly covered here, these are see-and-avoid/ sense-and-avoid and workload. In the area of see-and-avoid, it was noted that most UAS aircraft have not been designed for visual conspicuity. Improvements in this area can be made through high visibility colors and through the use of strobe and/or anti-collision lights. It was noted that the Light Emitting Diode (LED) strobes can even be used on small UAS. With regard to sense-and-avoid, answers to many of the questions indicate the desire of both ATC and manned aircraft pilots to know the presence of the UAS (such as through ADS-B), so advisories can be issued if needed by ATC, or for pilots, whether the UAS will respond to a TCAS RA. The UAS pilots also noted that the mission for UAS is typically quite different from that of manned aircraft in that it is typically not a Point A to Point B operation, and may involve sustained operations in a certain area with transits in and out to return to base.

UAS in the NAS have workload implications for all three groups, ATC, manned aircraft pilots and pilots of the unmanned aircraft. For the UAS pilot, there can be less workload than for a manned aircraft pilot if inner loop control is done by the aircraft (e.g., airspeed and altitude hold and fly heading). However, if failures occur, such as a global positioning system failure, high workload can occur as there may be no backup for the primary system. As noted in the survey responses, UAS camera imagery, as it exists at present, may not be of a resolution or field-of-view to assist in the piloting task. This seems an area ripe for research and development, especially in light of small low cost video sensors and on-board video processing to reduce downlink bandwidth. It has also been noted that GCS are typically not limited in terms of display area, so that has led to separate displays for different functions instead of intelligent integration of information which can reduce workload.

From the ATC perspective, it was noted that for military mixed operations of UAS and manned aircraft, an increased buffer is often needed around the UAS due to factors such as longer runway occupancy times or wake considerations following larger manned aircraft. For a controller used to the pacing of manned aircraft only operations, higher workload can result as additional traffic maneuvering may be required to establish and maintain the larger buffers. This higher workload may be evident especially for controllers new to this environment. It was reported that having a manned aircraft in the mix with UAS can actually result in lower ATC workload than a stream of UAS only, as the manned aircraft can respond and maneuver more quickly as well as self-separate from other traffic.

References

McCarley, J. S., and Wickens, C. D., (2005). Human Factors Implications of UAVs in the National Airspace. Technical Report AHFD-05-05/FAA-05-01, Federal Aviation Administration, Atlantic City, NJ.

U. S. Department of Defense, (2012). Unmanned Aircraft Systems (UAS) Ground Control Station Human-Machine Interface (HMI) Development & Standardization Guide. DoD Unmanned Aircraft Systems Task Force, Public Release 12-S-2388, V. 1, July 2012.

U. S. Government Accountability Office (2012). Unmanned Aircraft Systems: Measuring Progress and Addressing Potential Privacy Concerns Would Facilitate Integration into the National Airspace System. GAO-12-981, September 2012.

Verstynen, H. A., Foggia, J. R., Hoffler, K. D., (2010). An R&D Roadmap of UAS Access to the Next Generation Air Transportation System, Vol 1 (NASA ARD). NASA Langley TEAMS Contract NNL07AA00B, Task Order NNL10AM00T, December 17, 2010.

ISSUES RELEVANT FOR SYNTHETIC TEAMMATE – HUMAN TEAMMATE INTERACTIONS IN OPERATIONS OF A SYNTHETIC UNMANNED AERIAL SYSTEM

Nancy J. Cooke
Arizona State University and Cognitive Engineering Research Laboratory
Mesa, AZ
Christopher W. Myers
Air Force Research Laboratory
Dayton, OH

Two experiments are reported that set the stage for a project in which an ACT-R based Air Vehicle Operator will interact with two human teammates in an Unmanned Aerial System synthetic environment. Of interest are the ways in which the synthetic teammate fails to coordinate in an effective manner with humans. In Experiment 1, the new communication mode of text chat is compared to voice communications used previously in this task environment with all human participants. Issues of team performance and coordination were examined and differences noted particularly due to lag in the asynchronous chat mode. In Experiment 2 a condition in which two team members were told that the third was remotely located was compared to a condition in which two team members were told that the third human team member was a synthetic agent. Preliminary observations indicate that the "synthetic agent" is ignored and experiences terse communication compared to the remote teammate.

The research presented here is taking place as part of a larger project with the Air Force Research Laboratory that replaces a human UAS (Unmanned Aerial System) pilot with a cognitively plausible computational model that serves as a full-fledged synthetic teammate for a three-agent UAS ground control crew. Not only is the extension of the ACT-R cognitive modeling architecture of interest (Ball, Myers, Heiberg, et al., 2010), but the larger project will address questions about team coordination: What is the nature of coordination and collaboration (within all-human or mixed human-synthetic teams) in UAS ground control settings and what do deficiencies in synthetic teammate interactions with human teammates reveal about human-automation coordination needs?

Prior to inserting the synthetic teammate into the loop with two human participants, two experiments with all-human teams were conducted to establish baselines and are reported here. First, to establish a baseline for a new text chat mode of communication, team performance and coordination is examined using text chat communication and compared to voice communication. Second, performance and coordination in three-person teams was investigated when either two teammates were told that the pilot was remotely located or when they were told that the pilot was a synthetic teammate.

Team Coordination in the CERTT UAS-STE

This research program is conducted in the context of the CERTT UAS-STE (Cognitive Engineering Research on Team Tasks Unmanned Aerial System - Synthetic Task Environment (Cooke & Shope, 2005). The UAS-STE is based on the United States Air Force Predator UAS ground control station. The UAS-STE task requires a team of three people to complete the task of photographing critical waypoints. Each team member is assigned to one of three roles: an Air Vehicle Operator (AVO), a Payload Operator (PLO), or a Data Exploitation Mission Planning Coordinator (DEMPC). The DEMPC plans a mission route through multiple waypoints, the AVO is responsible for flying the simulated UAS and monitoring UAS systems, and the PLO takes photographs of designated waypoints and monitors camera systems. The roles are interdependent, where each role requires input from other team members to complete the team's goal of photographing designated waypoints. Further, the CERTT UAS-STE is dynamic and taking good photographs of designated waypoints requires information to be shared among teammates in a timely manner. A single UAS-STE mission consists of 11-12 targets and lasts a maximum of 40 minutes; each team performs five 40-minute missions.

Over a decade of research conducted in the CERTT UAS –STE has indicated that team interaction in the form of coordinated information passing and communication is important for predicting team performance and has led to a theory of Interactive Team Cognition (Cooke, Gorman, Myers, & Duran, 2012). In particular, coordination is based on the timely sending and receiving of information required for taking good photographs of designated waypoints. A coordination score (κ) is based on the timing and sequence with which key pieces of information are communicated among teammates (Gorman, Amazeen, & Cooke, 2010). The coordination score (κ) is computed as the amount of time from when information *I* about waypoint *w* is passed from the DEMPC to the team to when feedback *F* about taking a good photograph of waypoint *w* is provided to the team from the PLO. This is then divided by the amount of time from when the PLO and AVO negotiate *N* UAV flight dynamics for waypoint *w* up to when the PLO provides feedback *F* that a good photograph was taken for the waypoint *w*.

$$\kappa = \frac{F_w - I_w}{F_w - N_w}$$

For this project we hypothesized that some subtle timing and coordination behaviors would be absent in the synthetic teammate and would therefore impede coordination and effective performance. These two studies provide some additional baseline information about all-human coordination that will be used to test the synthetic teammate.

Coordinating with Synthetic Teammates

Prior to this project, team communication in the UAS-STE occurred via voice with microphones, headsets and push-to-talk intercom buttons. To avoid speech recognition issues and to transition to the text chat form of communications used commonly in the military, the synthetic teammate and the two human teammates will communicate via text chat. Because text chat is not a transient signal like voice and because communications can occur asynchronously, there is a possibility that coordination among teammates using text chat will be altered. Specifically the coordination score should be impacted by the asynchronous nature of communication. It is unclear whether performance will be affected, but if coordination is made more difficult, performance is also likely to be negatively impacted in this task. Not only will this experiment address questions about coordination and text chat, but will also provide a baseline against which to compare future performance and coordination data when the synthetic teammate is part of a team.

Team performance and coordination can also be affected by human teammates' expectations of the situation and of each other. Experiment 2 was conducted to ascertain how teammates would alter their interactions when they believed that the third teammate was a "synthetic agent". A number of individual behaviors were identified from past data that supported team coordination and these were also noted in Experiment 2.

Experiment 1: Text Chat vs. Voice Communications in the UAS-STE

The purpose of the first study was to collect baseline data in the context of the CERTT UAS-STE task with all human teams communicating via text chat, the mode of communication that will be used with the synthetic teammate. This mode was compared to voice communications, used in previous studies. Also, given the preponderance of text-based communications in our society and its adoption in time critical military and civilian contexts, the comparison of text versus voice as modes of communication is relevant and of increasing importance. By many accounts (Baltes, Dickson, Sherman, Bauer, and LaGanke (2002), Weeks, Kelly, & Chapanis (1974)), the use of test chat may not be the best mode of communication in time-pressured circumstances. The purpose of the experiment was to investigate how text-based communications affect team performance and coordination within the UAS-STE. Based on previous research, we hypothesized that teams communicating with text would coordinate differently from teams communicating using voice and that teams communicating with voice would perform the task better than those using text.

Method

Participants. Twenty, three person teams comprised of college students and the general population of the Mesa, Arizona area voluntarily participated in one 6.5 hour session. Individuals were compensated for their participation by payment of $10.00 per hour with each of the three team-members on the highest performing team receiving a $100.00 bonus. The majority of the participants were males, representing 75.9% of the sample. Individuals were randomly assigned to either a *voice* or *text* chat communication condition. The participants were also randomly assigned to teams and to one of three roles. All members of teams were unfamiliar with each other when they arrived for their sessions.

Equipment and Materials. The experiment took place in the CERTT Laboratory configured for the UAS-STE (described earlier). Participants in the TC condition communicated using the keyboard and a custom-built text communications system designed to log speaker identity and time information. The text communications interface was divided into 3 separate 'modules.' The 'receiver module' alerted participants with a lighted button when a message from another team member was sent. The receiver module also allowed participants to read incoming messages by pressing and holding the F10 key. On releasing the F10 key, the message was then displayed in the 'storage module,' which was comprised of a window that contained previously received messages in a list. Participants were given the ability to scroll through the messages by pressing the F7 and F8 keys. Participants sent messages with the 'transmit module.' To send messages, participants first typed their message in the transmit module window, selected the recipient using the F3, F4, and F5 keys, and then pressed F1 to send. The interface enabled participants to select multiple recipients. Each message was time stamped with when it was sent (F1 key-presses) and when it was received (F10 key-presses) in order to compute coordination scores (κ) and dynamics. Participants in the Voice Communications condition communicated with each other and the experimenter using David Clark headsets and a custom-built intercom system designed to log speaker identity and time information. The intercom enabled participants to select one or more listeners by pressing push-to-talk buttons.

Custom software (seven applications connected over a local area network) ran the synthetic task and collected values of various parameters that were used as input by performance scoring software. A series of tutorials were designed in PowerPoint for training the three team members. Custom software was also developed to conduct tests on information in PowerPoint tutorials, to collect individual taskwork relatedness ratings, to collect NASA TLX and SART ratings, to administer knowledge questions, and to collect demographic and preference data at the time of debriefing. This report will focus on performance and coordination data.

Procedure. The experiment consisted of one 7-hour session (see Table 1). The AVO was located in a separate room adjacent to the other members (DEMPC and PLO). The AVO entered the building through a separate entrance located on the opposite side of the building, and was not allowed to have contact with the other members until debriefing. In the session, the team members were seated at their workstations where they signed a consent form, were given a brief overview of the study and started training on the task.

The number of targets varied from mission to mission in accordance with the introduction of situation awareness roadblocks at set times within each mission. Missions were completed either at the end of a 40-minute interval or when team members believed that the mission goals had been completed. Following each mission, participants were given the opportunity to view their team score, their own individual score, and the individual scores of their teammates. The performance scores were displayed on each participant's computer and shown in comparison to the mean scores achieved by all other teams (or roles) who had participated in the experiment up to that point

Results

Team Performance. Team performance was measured using a composite score based on the result of mission variables including time each individual spent in an alarm state, time each individual spent in a warning state, rate with which critical waypoints were acquired, and the rate with which targets were successfully photographed.

Penalty points for each of these components were weighted *a priori* in accord with importance to the task and subtracted from a maximum score of 1000. Team performance data were collected for each of the five missions.

Team performance was analyzed using a 2 (text, voice) x 4 (mission) mixed ANOVA. Each communication condition (text, voice) had 10 teams. There was a main effect of mission $F(3, 54) = 9.447$, $p < .001$. Teams improved their performance score across the first four missions. There were no significant effects of communication condition, $F(1, 18) = 0.57$, $p < 0.46$, although the voice teams consistently had higher performance scores across all missions than teams in the text chat condition (see Figure 1).

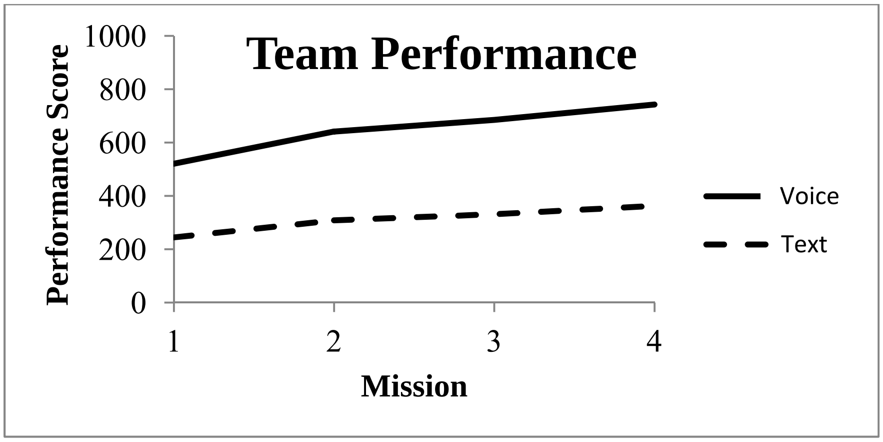


Figure 1. Team performance means for each mission differed over missions, but not condition.

LSD pair-wise comparisons showed that team performance improved over the course of the first four missions, with significant gains between the first two missions ($p = .005$) and between the second and fourth missions ($p = .015$).

Coordination. Based on the inherent time costs of using text chat (e.g., typing, noticing a message arrived, etc.), there was a significant time lag between when a message was sent and when it was received (M = 10.5 s for text; 0 s for voice). To determine if there was a difference in coordination score between voice and text chat, a 2 (communication mode) x 4 (lower workload missions) mixed ANOVA was conducted on coordination scores. There was a significant main effect for which text chat had a significantly lower coordination score than voice (p = 0.042). This is not to say that the voice condition coordinated "better," but only to say that the two communication conditions coordinated differently. Further, a measure that reveals the stability of team coordination dynamics, the Hurst exponent, was also analyzed to determine if there was a coordination stability difference between communication groups. An independent samples *t*-test on the average Hurst exponents across teams revealed that text chat teams were, on average, coordinated in a more stable fashion ($M = 0.9527$, $SD = 0.0131$) than voice teams ($M = 0.8988$, $SD = 0.061$), $t(15) = 2.287$, $p = 0.037$.

For the four low workload missions the median of the performance scores was 310 in the chat condition, with 5 teams below the median and 5 teams above the median. A regression analysis on all the teams combined revealed that the linear trend between communication lag and team performance was significant ($F\ (1, 38) = 9.06$; $p = 0.005$) indicating that as lag decreased, performance increased. Regression analyses also revealed a positive linear relationship between performance score and Kappa in teams performing above the median performance score ($F\ (1, 13) = 4.46$; $p = 0.055$). Overall these results indicate that text chat results in different coordination patterns than voice chat and that there is a relationship between these patterns and team performance.

Experiment 2: Expecting a Synthetic Teammate

In this particular study, we examine how teammate interactions (via text chat) are affected by expectations that the pilot is either a synthetic agent or a human teammate. Three person teams were arranged so that the pilot station and pilot were not visible to the other two teammates (mission planner and photographer) and half of the teams were informed that the pilot was a synthetic agent and half were informed that the pilot was a remotely-located human teammate. However, in both conditions the pilot was a human participant. Measures of individual and team performance, coordination, team process, team situation awareness and knowledge were collected over four 40-minute missions. We predict that the expectation that the two teammates are interacting with a human or synthetic pilot will alter coordination and communication patterns. These results will inform design of the synthetic agent and provide baseline data for a Turing-like test of synthetic teammate validity in the next experiment.

Method

Participants. Twenty, three person teams comprised of college students and the general population of the Mesa, Arizona area voluntarily participated in one 6.5 hour session. Individuals were compensated for their participation by payment of $10.00 per hour. Individuals were randomly assigned to one of two conditions: remote AVO or synthetic AVO. The participants were also randomly assigned to teams and to one of three roles. All members of teams were unfamiliar with each other when they arrived for their sessions.

Equipment and Materials. Participants all used the custom text chat capabilities as in the previous study except the interface was slightly improved to make it easier to use. All other equipment and materials were identical to Experiment 1.

Measures. For the purpose of this experiment the measures used in Experiment 1 were also used in Experiment 2. In addition a performance score was calculated for each target based on the timely and accurate processing of a target. In addition a set of behaviors related to team coordination were identified in previous data sets and were noted whenever they occurred in this study. The behaviors are listed in Table 1.

Table 1.
Individual Behaviors Supportive of Team Coordination

Negative Communication
- Argue – *DEMPC and AVO can argue over the best way to give upcoming waypoint restrictions?*
- Specific to chat conditions
    - Timing – AVO sends text asking for next waypoint just as DEMPC texts the next waypoint info.
    - Lag in response – PLO asks a questions that is not answered until multiple unrelated texts have been posted.

Positive communication
- Help out – *PLO tells DEMPC, "Please give info next target info to AVO."*
- Acknowledge members' speech – *"Roger that."*
- Give praise – *Good job guys!*
- Check with others before implementing a decision – *PLO asks AVO, "I am about to take a pic, are we at 2000 feet?"*
- Clarification – AVO asks DEMPC to clarify what was meant in a previous message.

Repeated Requests
- Same info or action requested two or more times
- PLO asks repeatedly for information needed to take a photo.

Unclear Communications
- Misspellings, ambiguous terms, experimenter cannot understand

General Status Update
- Inform others of current status – AVO tells PLO "I am at 2500 feet now."

Inquiry About Status of Others
- Inquire about current status of others – *DEMPC asks AVO "How are we doing on our heading/fuel etc."*
- Express concern – *DEMPC asks AVO "Are we headed to the next target? We appear to be off course."*

Planning
- Anticipate next steps – *AVO asks DEMPC, "Where are we going after LVN?"*

Suggestions to Others

- Make suggestions to other members – *DEMPC tells AVO to increase speed in route to targets and slow down upon arrival.*

Procedure. Three person teams were arranged so that the pilot station and pilot were not visible to the other two teammates (mission planner and photographer) and half of the teams were informed that the pilot was a synthetic agent and half were informed that the pilot was a remotely-located human teammate. However, in both conditions the pilot was a human participant. Measures of individual and team performance, coordination, team process, team situation awareness and knowledge were collected over four 40-minute missions. In all other respects the procedure was like that of Experiment 1.

Preliminary Results

Data collection is ongoing, but preliminary observations of experimenters indicate that in the synthetic teammate condition, the human participants tend to ignore the synthetic teammate and speak tersely, eliminating any of the social or polite discourse that typically occurs with a human teammate.

Conclusion

In these studies we have set the stage for synthetic teammate validation by 1) collecting baseline data in the UAS-STE using text chat, 2) by identifying individual behaviors that support coordination, and 3) by understanding human teammate predilections to interact differently with a synthetic teammate vs. a human teammate.

References

Ball, J. Myers, C., Heiberg, A. Cooke, N. J., Matessa, M., Freiman, M., & Rodgers, S. (2010). The synthetic teammate project. *Computational and Mathematical Organization Theory*, *16,* 271-299. DOI 10.1007/s10588-010-9065-3

Baltes, B. B., Dickson, M. W., Sherman, M. P., Bauer, C. C., & LaGanke, J. S. (2002). Computer-mediated communication and group decision making: A meta-analysis. *Organizational Behavior and Human Decision Processes, 87*(1), 156-179.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J.L. (2012). Interactive Team Cognition, *Cognitive Science, 1-31,* DOI: 10.1111/cogs.12009.

Cooke, N. J., & Shope, S. M. (2005). Synthetic Task Environments for Teams: CERTT's UAV-STE. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas & H. Hendrick (Eds.), *Handbook of Human Factors and Ergonomics Methods* (pp. 46-41 - 46-46). Boca Raton, FL: CRC Press.

Gorman, J. C., Amazeen, P. G., & Cooke, N. J. (2010). Team Coordination Dynamics. *Nonlinear Dynamics Psychology and Life Sciences, 14,* 265-289.

Weeks, G. D., Kelly, M. J., & Chapanis, A. (1974). Studies in interactive communication: V. Cooperative problem solving by skilled and unskilled typists in a teletypewriter mode. *Journal of Applied Psychology, 59*(6), 665-674.

Acknowledgements

This work is supported by ONR N000140910201 and N00014-11-1-0844 to the Cognitive Engineering Research Institute. The authors would like to thank Jasmine Duran, Jamie Gorman, Harry Pedersen, and F. Erik Robinson for their contributions to the first study.

THE INTEGRATION OF NEW TECHNOLOGY INTO A COMPLEX SYSTEM

Kelly Neville, Beth Blickensderfer, Katherine Kaste, Stephen L. Dorton
Embry-Riddle Aeronautical University
Daytona Beach, FL

Stephen P. Luxion
U.S. Air Force Reserve Officer Training Corps
Embry-Riddle Aeronautical University
Daytona Beach, FL

This paper describes an evaluation of the impacts of introducing change into the established complex system of UAS operations. Two technologies not currently used in UAS operations, a backup communications system and a traffic display, were operated by Guardian UAS pilots as they shadowed live UAS flights in a back-up control station. The flights were demonstration rather than research flights; we nonetheless were able to make the most of the opportunity and collect observation, survey, and interview data to gain insight into effects of the technology insertions. Technology-insertion impacts and recommended technology adaptations were categorized into emergent themes. The identified themes were found to align with three of the five basic generic support requirements for cognitive work proposed by Woods (2005, Generic support requirements for cognitive work: Laws that govern cognitive work in action. *Proceedings of the HFES 49th Annual Meeting*. Santa Monica, CA: HFES).

When changes are made to a complex system, effects can be difficult, if not impossible, to predict. Analysis and modeling are not likely to predict the effects of a change to a truly complex system. The best evaluation strategy may instead be to insert a change on a trial basis. Snowden and Boone (2007) describe this complex-systems strategy as inserting a *probe* so that the resulting emergent patterns can be assessed.

The main goal of the work described in this paper was to evaluate the effects of introducing into UAS operations two technologies, i.e., “probes”, with potential to facilitate the integration of UAS into the national airspace system (NAS). One technology was a backup voice communications radio system to be used if a UAS pilot loses the voice communications link with air traffic control (ATC). UAS pilot communications with ATC are relayed via the aircraft so when the datalink connection with the aircraft is lost, voice communications are also lost. The second technology was an air traffic display (see Figure 1), which would provide UAS pilots with their only visual source of traffic information.

Figure 1. The Garmin GMX 200 traffic display used in the technology demonstration.

The technologies were introduced into UAS operations for the purpose of a technology demonstration. UAS pilots interacted with the technologies during staged *NAS events* while they shadowed the unfolding of live missions flown by two colleagues. The naturalistic setting helped pilots to consider the new technologies in terms of their relationships with the existing UAS operations system and their support for mission activities within the NAS.

Methods

Participants

Two UAS pilots stationed at Cape Canaveral Air Station (CCAS) volunteered to participate in the technology demonstration. Both read and signed informed consent documents. One pilot (*Pilot A*) had approximately 4,500 hrs as a UAS pilot and close to 2,000 hrs in general aviation aircraft. The other (*Pilot B*) had accrued approximately 500 hrs as a UAS pilot and 3,700 hrs flying military aircraft.

The pilots took turns as the *pseudo pilot* during each of three UAS missions flown by a pilot and sensor operator in a nearby ground control station (GCS). As pseudo pilot, the pilot sat in the operations center where he could monitor GCS radio communications and observe mission progress on displays identical to those in the GCS. One 2- to 3-hr mission was flown on each of three sequential days.

Technology Additions

The backup communications system prototype was developed for the demonstration by Harris Corporation. The prototype consisted of a headset, user interface with voice-to-digital signal conversion, Jotron AM radio, and portable communications tower. The prototype's user interface featured a small screen, approximately 6 x 4 in (15.2 x 10.2 cm), positioned over a row of five function buttons and flanked on either side by system navigation controls.

The 4 x 3 in (10.2 x 7.6 cm) Garmin GMX 200 traffic display was attached to the upper right corner of a desktop monitor. The display featured icon representations of the UAS and surrounding traffic, directional information, and range rings (see Figure 1). Displayed aircraft 'tracks' were derived from ATC surveillance radar and transmitted via the Traffic Information Services-Broadcast (TIS-B) Service.

A desktop monitor to which the Garmin GMX 200 was attached displayed air traffic over a large section of the southeast United States. Evaluation team members used this display to gauge the accuracy of traffic data shown on the Garmin traffic display. Its position in front of the pseudo pilot may have interfered with assessing the Garmin traffic display, although data do not point to this.

Procedure

The Guardian variant of the MQ-9 UAS (General Atomics Aeronautical Systems) flew a saw-toothed flight path off the central east coast of Florida (see Figure 2) on each of three demonstration flights. Altitude was maintained at 20,000 ft between the initial climb and final descent. The pilot in the GCS was asked to fly the prescribed route and was not given other taskings. Study participants assessed the mission plan as much simpler and easier than their typical mission.

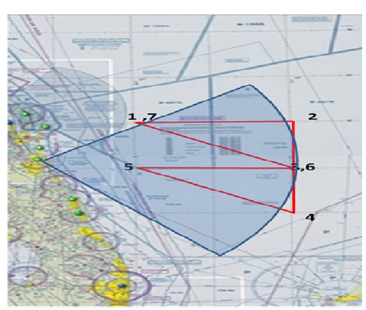


Figure 2. The mission flight path flown in the demonstration, indicated by red lines.

Pseudo pilots were seated in front of interfaces to the two new technologies. Immediately before the start of the first mission, pseudo pilots were familiarized with the technologies and shown how to use them. During each mission, two preplanned NAS events were presented to the pseudo pilots at Waypoint 3 and at three additional points spaced 10 min apart. The two events were always the same and represented events that could occur in the NAS. The pseudo pilots were instructed to imagine they were flying in the NAS when the events occurred.

In the first NAS event, pseudo pilots were told that the communication link with ATC was lost. The pseudo pilots used the backup communications system to re-establish communications, a task that culminated with a radio check to the local CCAS control tower. In the second NAS event, a veteran air traffic controller on the evaluation team issued a traffic call for traffic coming within 10 miles of the aircraft. (Distant live traffic was made to appear as if it was within 10 miles by treating range ring distances as if they represented smaller-than-actual distances.) The pseudo pilots used the Garmin CDTI to locate traffic and issued a verbal acknowledgement when they sighted it on their display.

In the operations center, a researcher seated behind the pseudo pilot observed and took notes. Because the pseudo pilots were not actually flying an aircraft, their workload was low enough that they could volunteer verbal feedback about the technologies over the course of each mission and immediately after each set of NAS events. The researcher captured this feedback in his or her notes.

After the first and third missions (participants were unavailable after the second mission), pseudo pilots completed a 20-item questionnaire and participated in a 1-hr semi-structured individual interview. The questionnaire consisted of 7-point-rating-scale and short-answer questions about the difficulty of the mission, ease and benefit of using the communications system, helpfulness and workload of the traffic display, and whether either technology changed their work. Interview questions had similar foci and asked pilots to describe their interactions with the technologies. The second interview, conducted after the third mission, revisited the questions asked after the first mission and then moved on to questions about past UAS missions for a different project. Interviews were tape recorded and subsequently transcribed.

Results and Discussion

Observer notes, questionnaire responses, and interview transcripts were assessed to gauge effects of the inserted technologies on UAS operations. Data are summarized below in terms of benefits and recommended adaptations:

- *Benefits* – Benefits the pilots experienced or anticipated the technologies could offer.
- *Recommended adaptations* – Changes that should be made so the technologies better integrate with and support UAS operations.

In addition, recommended adaptations to the traffic display were coded using five *generic requirements for systems that support cognitive work* proposed by Woods (2005) and Elm, Potter, Tittle, Woods, Grossman, and Patterson (2005). This *thematic analysis* (Braun & Clarke, 2006) revealed patterns in traffic display adaptations that should improve the display's integration into the complex cognitive UAS operations system. It additionally served as a check on the adequacy of the five generic requirements specified by Woods and his colleagues.

Backup Communications System Effects on UAS Operations

Both pseudo pilots gave positive ratings (6s and 7s on the 7-point scale) of the backup radio system. They rated the system as easy to use and beneficial to the recovery of communications with ATC. Pilot A described his use of the system as "pretty seamless" and referred to it as the "most functionally promising" of the inserted technologies. Pilot B was similarly positive in his comments (e.g.: "if [the system] works as advertised, it will greatly increase our comms capability with ATC.") At the same time, the pilots were concerned that the technology might be dropped into their operations without first being integrated with their work or existing systems.

Benefits. Currently, a lost communications link means a UAS crew must call the ATC center via telephone, and someone other than the sector controller will answer that call. The pilots indicated that the new capability fills a gap by allowing them to re-establish communications with the sector controller within seconds. Pilot A suggested that the technology has the capacity to "expand our operations envelope and marginalize our risk…with proper integration and development."

Recommended adaptations. Noting that stress will be high during events that require the backup technology, Pilot A pointed out that "In order for us to pick up another headset and tune in another frequency…it's only going to add workload at a time when we're lost-comms and need all the efficiency in our HMI [human-machine interface] as possible." On the other hand, "if…channel selection… emulates the simplicity of the test, then that would be ideal." In other words, the pilots want the technology to be evaluated under more demanding conditions and in conditions for which the technology was not preconfigured. (The communications system was preconfigured for the demonstration flights.)

Traffic Display Effects on UAS Operations

The pseudo pilots gave positive ratings (5 to 7 on a 7-point scale) of the traffic display's helpfulness and benefit to situation awareness (SA). Pilot workload associated with using the traffic display during a mission was rated as 4 and 5 ('1' signifies low workload) for the first mission. For the third mission, Pilot A gave a midrange rating of 4 and Pilot B came down to a 2. Pilot B attributed workload ratings to learning to manipulate a new piece of equipment (the CDTI). Pilot A attributed workload to a lack of integration with other displays.

Benefits. Both pilots viewed the traffic display as useful for maintaining awareness of their aircraft's position relative to its surroundings and planning ahead. Pilot B additionally said he would use the traffic display to communicate UAS position and path to ATC as part of a lost-link procedure. Pilots described using the display to obtain "overall SA" and for "response to and identification of ATC-directed targets" (Pilot A) and as a means to "enhance…overall situational awareness" (Pilot B).

Recommended adaptations. Responses to inserted events revealed that communications protocols associated with use of the traffic display need attention. The current protocol that pilots use to reply to traffic calls assumes the pilot can see traffic directly, versus only as an icon on a display. New phraseology should additionally address a concern voiced by Pilot B; specifically, how to establish that a track the pilot is "seeing" and confirming is the same track ATC called out.

Whereas the pilots found the traffic display to enhance SA and "big-picture awareness" of the Command Duty Officer (CDO; oversees and coordinates UAS missions), they reported it to be of limited value for tactical, close-in traffic avoidance by pilots. Pilot B stated the display would be valuable to him in the role of CDO, but that "in the GCS, I've got enough screens and buttons to push. I've got enough to do. It's something I wouldn't do. I wouldn't be interested in it when I'm flying."

Thematic Analysis of Recommended Traffic Display Adaptations

Pilot interview data and short-answer survey responses were synthesized into 16 specific recommendations for traffic display adaptation. These recommendations were coded using the five generic requirements for system designs that support cognitive work and examples Woods, Elm and their colleagues use to define each code (Elm et al., 2005; Woods, 2005). Codes and the number of display adaptation recommendations assigned to each are shown in Table 1.

Table 1.
The Number of Recommended Traffic Display Adaptations Assigned to Cognitive-Work-Support Codes

Generic Cognitive Work Support Codes		**Number of Recommendations**
Generic Requirements	**Requirement helps the operators in a system to…**	
Observability	gain insight into system processes.	1
	avoid keyhole effect.	1
	see sequences.	2
	see future activities and contingencies.	
	see patterns and relationships in a process.	3
Directability	direct/re-direct resources in response to and anticipation of changes in the environment.	3
	direct/re-direct activities.	
	direct/re-direct priorities.	1
Teamwork with Agents; Shifting Perspectives	establish teamwork (with human and automated teammates)	
	coordinate and synchronize activity across agents.	
	redirect teammates by seeding new ideas, reminding, and critiquing as a situation changes.	
Directed Attention	reorient attention in a changing environment.	3
	track teammates' focus of attention.	
	judge interruptibility of teammates.	
Resilience	use failure-sensitive strategies (via feedback).	
	explore outside current boundaries or priorities.	
	step in to support brittle automation.	
	maintain peripheral awareness.	
Other	interpret symbols and alarms.	2

Patterns in Recommended Adaptations. Pilots' recommendations for adapting the traffic display mainly involved supporting perception and cognition in a dynamic information-rich environment. The pilots wanted information presented in ways that support a majority of the *observability* goals (i.e., ways observability supports complex cognitive work) listed in the first 5 rows of Column 2 in Table 1. They also wanted to have continuous control over the information they viewed; specifically, they wanted to be able to visually segregate, flag, and adapt displayed information. The third main category of recommendations involved changes to improve pilots' ability to perceive and process important traffic information without removing their attention from other aspects of their work.

Implications for the Generic Requirements for Supporting Cognitive Work. We were able to map all traffic display recommendations to the generic system requirements for supporting cognitive work with just two exceptions. Those two exceptions were related to improving the interpretability of symbols and alarms (see the last row of Table 1) and are relevant to the observability requirement in that they make observability goals possible. In addition, three recommendations that mapped to the generic requirement *directability* did not clearly map at the goal level. These recommendations were coded as *direct/re-direct resources*; however, the resources in these recommendations were traffic display information elements, which are not typically considered work-system resources. This category of recommendation can continue to be treated as a resource or it might suggest the need for an additional directability goal: the goal of helping the operator in a system *direct/re-direct the form and content of a system's human-technology interfaces*.

Conclusions

Adding new technology to a complex system can introduce perturbations and interactions that, in effect, negate the technology's intended benefits. By revealing ways to align safety-enhancing technologies with the UAS operations system, including pilots' cognitive work, the demonstration described in this paper contributes to a reduced risk of unintended consequences.

According to complexity science, a system develops resilience when its parts are permitted to co-evolve (e.g., Bar-Yam, 2004; Benbya & McElvey, 2006). In this effort, UAS pilots were given new technologies to use while responding to inserted NAS events as they shadowed live UAS operations. By combining these system elements (pilots, new and existing technologies, NAS events, and UAS mission activities), we gained insight into ways they can co-evolve and become more attuned with each other.

This work contributes to the study and design of complex systems by evaluating five proposed generic requirements for systems that support cognitive work (Elm et al., 2005; Woods, 2005). Support was found for three of the five requirements and no evidence was found to conflict with any of the five. Findings suggest that information interpretability might be added as a sixth requirement, although that requirement might be better suited to the body of traditional human-computer interface design guidelines than to Woods' and Elm et al.'s five generic requirements. Data also suggested that the directability requirement be extended to include directability of the human interfaces to system elements and activities in addition to directablity of system elements, activities, and priorities.

Acknowledgements

This work was funded by the FAA Advanced Technology Development and Prototyping Group. Opinions and interpretations presented in this paper are those of the authors and do not represent official positions or policies of the FAA or Embry-Riddle Aeronautical University (ERAU). Special thanks to the CBP at CCAS and the technology demonstration team, led by Damon Thomas (FAA), Todd Waller (ERAU), Keven McEntee (Saab-Sensis), and Dan Berlinrut (JMA Solutions). We are also grateful to Nancy Cooke (Arizona State University), Drew Gellerson (CBP), Kevin McEntee, Ben Walsh (FAA), Erin McCollum (Sikorsky), and Tim Buker (SAIC) for commenting on earlier versions of this paper.

References

Bar-Yam, Y. (2004). *Making things work. Solving complex problems in a complex world.* Cambridge, MA: NECSI Knowledge Press.

Benbya, H. & McKelvey, B. (2006). Using coevolutionary and complexity theories to improve IS alignment: A multi-level approach. *Journal of Information Technology, 21*, 284-298.

Braun, V. & Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology, 3(2),* 77–101.

Boone, M.E. & Snowden, D.J. (2007). A leader's framework for decision making. *Harvard Business Review, November 2007,* 69-76.

Elm, W., Potter, S., Tittle, J., Woods, D., Grossman, J., & Patterson, E. (2005). Finding decision support requirements for effective intelligence analysis tools. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*. Thousand Oaks, CA: Sage Publishing.

Woods, D.D. (2005). Generic support requirements for cognitive work: Laws that govern cognitive work in action. *Proceedings of the HFES 49th Annual Meeting*. Thousand Oaks, CA: Sage Publishing.

COORDINATION IN DISTRIBUTED UNMANNED AIRCRAFT SYSTEMS

Rosemarie Reynolds
Embry Riddle Aeronautical University
Daytona Beach, Florida
C. Shawn Burke
Institute for Simulation and Training
Orlando, Florida
Eduardo Salas
Institute for Simulation and Training
Orlando, Florida

Our objectives in this paper are to: (1) explicate the link between UAS accidents and coordination, and (2) to provide prescriptive guidelines for improving coordination in distributed UAS. To accomplish these goals we will review the literature on team coordination, examine how distribution impacts that process, identify the areas in which coordination plays a critical role in UAS incidents and accidents, and conclude with some prescriptions for improving UAS team coordination.

Unmanned aircraft systems (UAS) have been in use since 1967, when the Israeli Army used radio-controlled model aircraft to take aerial photographs behind Egyptian lines (Nelson & Bolia, 2006). Although still primarily used in military applications, there are a great many potential uses for UAS in the civilian sector, including aerial surveillance, oil, gas and mineral exploration, logistics, scientific research, and search and rescue (Adams, Humphrey, Goodrich, Cooper, Morse, Engh, & Rasmussen, 2009; DeGarmo, 2004; Xiang & Tian, 2011). Despite the growing acceptance of UAS, a cause for concern is the current UAS accident rate, which is sometimes estimated to be 100 times that of manned aircraft (Hing & Hing, 2009). For example, the U.S. Air Force recorded six serious UAS accidents (resulting in a fatality, permanent disabling injury, or property damage of $1 million or more) per 100,000 flight hours in fiscal year 2007 and 2008. Air Force accident rates for crewed aircraft in 2009 and 2010 were 0.72 and 0.80 per 100,000 flight hours respectively (Goyer, 2010; Nullmeyer, 2009). For all military UAS combined, the Office of the Under Secretary of Defense (2004) reported the loss rate for the Predator at 32, the Pioneer at 334, and the Hunter at 55 (all for much less than 100,000 flight hours). For comparison purposes, the loss rate per 100,000 flight hours during the same period for the F-16 was three, for General Aviation it was one, and for large airlines it was 0.01. Similar figures hold true for non-military UAS as well; the accident rate for UAS flown by Customs and Border Protection (CBP) was recently reported as 52.7 accidents per 100,000 flight hours (Kalinowski, 2010).

As Lacher, Zeitlin, Maroney, Markin, Ludwig, and Boyd (2010) pointed out, “The fundamental difference between manned aviation and unmanned aviation is that the pilot is not physically on-board the unmanned aircraft (p.1).” This simple – and obvious – fact makes it easy to forget that while an UAS does not include a human aboard the aircraft, a UAS is, in fact, a complex system run by a team, whose members have clearly differentiated and interdependent tasks such as Air Vehicle Operator, Payload Operator, and Mission Planner. As a result, an UAS presents the usual coordination challegnes of teamwork, workload, and situation awareness. In addition, there are several human factor challenges unique to UAS. As McCarley and Wickens (2005) pointed out:

> UAV flight presents human factors challenges different from and in some ways greater than those of manned flight. These arise primarily from the fact that operator and aircraft are not co-located. As discussed in more detail below, the separation of operator and vehicle imposes a number of barriers to optimum human performance, including loss of sensory cues valuable for flight control, delays in control and communications loops, and difficulty in scanning the visual environment surrounding the vehicle. (p. 1)

The operator and the aircraft are not the only distributed components of the UAS system. Advances in technology have made it ever more likely that an UAS team - in both military and civilian applications - may be distributed, rather than co-located, as well. An extensive body of research has indicated that such distributed teams often have team processes, especially in the areas of communication and coordination that differ from those experienced by co located teams (Kiekel, Gorman, & Cooke, 2004, Rapp & Mathieu, 2007; Reynolds & Brannick, 2009; Schiller & Mandviwalla, 2007; Stone & Posey, 2008). We therefore suggest that team coordination may be one of the largest human factors concerns for a UAS system, and bolster our argument with a brief review of team coordination and how distribution impacts that process. We then provide evidence from the literature which suggests that coordination plays a critical role in UAS accidents. We conclude with some prescriptions for improving coordination in UAS teams.

Team Coordination

A team is a type of group characterized by differentiated tasks and high levels of interdependence (Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). This interdependence is what Bell and Kozlowski (2002) called *intensive interdependence*, in which team members must "diagnose, problem solve, and/or collaborate simultaneously as a team to accomplish their task (p 18-19)." Teamwork represents how *well* the team members coordinate the interdependent components of performance (Salas, Cooke, & Rosen, 2008). Coordination, or making sure that the right thing happens at the right time, may be one of the most critical aspects of teamwork.

Coordination in Distributed Teams

With reference to teams, *distributed* means geographically, organizationally, or temporally dispersed. Therefore a distributed team (DT) is a group of people with clearly specified roles working interdependently on a common goal, using use some form of computer technology to coordinate at least part of their work. This computer mediated communication (CMC) impacts all of the team processes. The focus during what Marks, Mathieu, and Zaccaro (2001) called the *Transition Process* is on planning, mission analysis, goal specification and strategy formulation. For a DT, a lack of shared context and shared knowledge makes this stage difficult.

During the *Action Process*, the focus is on goal accomplishment through activities such as monitoring progress, sharing knowledge, decision making, communication, and coordination (Marks, Mathieu, & Zaccaro, 2001). In a distributed team, the time delays in sending feedback, the lack of a common frame of reference, the lack of non-verbal feedback, and lack of trust all make coordination difficult (Driskell, Radtke, & Sales, 2003; Hambley, O'Neil, & Kline, 2007; Powell, Piccolo, & Ives 2004; Reynolds & Brannick, 2009).

`Why Coordination is an Issue in UAS

In 1996, the Air Force Scientific Advisory Board identified the human/system interface as the greatest deficiency in UAS designs (Williams, 2006). Other research also supports the idea that human factors are a major concern in UAS. Asim, Ehsan, and Rafique (2010) examined a sample of 56 UAS accidents, and found that human factors were a contributing cause in 32% of them, while Williams (2004), in an analysis of 74 accidents, found that the percentage of accidents attributable to human factors varied across platforms, with 21% of accidents with the Shadow attributable to human error, and 67% with the Predator. Manning, Rash, LeDuc, Noback, and McKeon (2004) investigated the role of human error in U.S. Army UAV accidents and found that human error was present in 32%.of the reported accidents, and, the authors noted that the U.S. Army attributed the lack of good crew coordination as a causal factor.

Crew coordination is made difficult in UAS due to limited bandwidth, and volume of information transmitted (Mouloua, Gilson, Kring, and Hancock, 2001; Van Erp, 2000). One area that appears to be critical in a UAS are the demands placed on coordination during task switching and handoffs between vehicles, payloads, missions, targets, and crew (Fern & Shively, 2011; McCarley & Wickens 2005). Similar findings were reported in a cognitive task analysis of UAS piloting, in which the authors noted that two major concerns were: 1) the competing demands for attention, and 2) a broken or compromised control loop, such as one caused by latency which affects pilot communications, or even a totally lost datalink (Neville, Blickensderfer, Luxion, Kaste, & Archer, 2012).

Possible Solutions to UAV Coordination Issues

Possible solutions for these coordination issues come from a variety of sources: 1) traditions human resource practices such as selection and training, 2) lessons learned from distributed team research, and 3) UAS technology developments.

Human Resource Management. Traditional human resource management practices such as task analysis and training represent one avenue for improving UAS team coordination. Neville, Blickensderfer, Luxion, Kaste, and Archer (2012), as a result of their cognitive task analysis, suggested that the boundaries and responsibilities in the relationship between UAS mission and piloting work need to be clarified, as conflict between mission priorities and piloting responsibilities was a concern. Adams, Humphrey, Goodrich, Cooper, Morse, Engh, and Rasmussen (2009) also used cognitive task analysis to clarify the roles and responsibilities in Wilderness search and rescue mini-UAVs, after having discovered that there were severe coordination deficiencies between the UAS and ground search resources.

One option for training is, of course, a modified form of Crew Resource Management (CRM). Sharma and Chakravarti (2005) noted that "It is prudent to plan CRM training for UAV operators as well. The sole aim of the training shall be to resolve the inherent problems of integrating a collection of technically proficient individuals into an effective team for all situations (p. 36)." Other methods of training could be useful as well. Salas, Nichols, and Driskell (2007) performed a meta analysis of team training interventions, and found a significant tendency for team training to lead to an increase in performance, with the most effective training being that which focused on coordination and adaptation.

Lessons learned from DT. Research indicates that process and performance in a distributed team is helped by a well-defined task structure, goal setting, clearly de ning roles and responsibilities, setting milestones, and agreeing on communication media and frequency (Gibson & Gibbs, 2006; Maynard, Mathieu, Rapp, & Gilson, 2012; Staples & Webster, 2008; Timmerman & Scott, 2006; Walther & Bunz, 2005). Technology focusing on teamwork rather than on flight may be of use as well; a well-designed user interface increases participation, trust, and cooperation, while easy to interpret awareness displays containing information about a remote collaborator's workload lead to communication attempts that are less disruptive, reduce workload, and increase team performance (Chang & Lim, 2006; Dabbish & Kraut, 2008; Strang, et al., 2011). Rusman, Bruggen, Cörvers, Sloep, and Koper (2009) evaluated the effectiveness of a standardized profile and found that the profile helped team members to form an impression of each other during the early stages of a team project.

Although DeLuca, Gasson, and Kock (2006) noted that members of DTs adapted their communication to low richness media by being more precise, concrete, concise, and complete, Belanger and Watson-Manheim (2006) found that DT members preferred using different media for different goals. This suggests that it may be helpful to vary the technology used by the team depending on the task; for example video conferencing during briefing and debriefing, synchronous communication during missions, and asynchronous when not on a mission.

UAS technology. As Williams (2006) pointed out, until recently, most displays have focused on flight, rather than coordination. Research focused on improving communication technology in UAS includes research on airspace display formats for both pilots and camera operators, as well as displays to reduce operator workload during handoffs (Draper, Geiselman, Lu, Roe, & Haas, 2000; Fern, & Shively, 2011). Van Breda, Jansen, and Veltman (2005) have been examining graphic overlays, ecological interface design, head-coupled control, and the use of prediction techniques to help compensate for image degradations. Dixon, Wickens, and Chang (2005) examined the effect of auditory autoalert and an autopilot on: (a) mission completion, (b) target search, and (c) systems monitoring, and found that both systems improved overall performance by reducing task interference and alleviating workload.

Conclusion

Coordination, or making sure that the right thing happens at the right time, may be one of the most critical aspects of teamwork. For a UAS, a number of factors make coordination difficult: 1) distribution, 2) lack of shared context and shared knowledge, 3) time delays in sending feedback, 4) lack of non-verbal feedback, 5) limited bandwidth, and 6) the sheer volume of information transmitted. These difficulties become particularly salient during high workload situations such as task switching and handoffs between vehicles, payloads, missions, targets, and crew. Possible solutions for these coordination issues come from a variety of sources: 1) traditions human resource practices, 2) lessons learned from distributed team research, and 3) UAS technology developments.

References

Adams, J. A., Humphrey, C. M., Goodrich, M. A., Cooper, J. L., Morse, B. S., Engh, C., & Rasmussen, N. (2009). Cognitive task analysis for developing unmanned aerial vehicle wilderness search support. *Journal of Cognitive Engineering and Decision Making, 3*(1), 1-26.

Asim, M., Ehsan, N., & Rafique, K. (2010). *Probable causal factors in UAV accidents based on human factor analysis and classification system.* Paper presented at the 27th Congress of International Council of the Aeronautical Sciences, Nice, France**.**

Belanger, F., Watson-Manheim, M.B. (2006). Virtual Teams and multiple media: Structuring media use to attain strategic goals. *Group Decision and Negotiation, 15* (4), 299-321.

Bell, B., & Kozlowski, S. J. (2002). A typology of virtual teams: Implications for effective leadership. *Group & Organization Management, 27* (1), 14-49.

Chang, K. T., & Lim, J. (2006). The role of interface elements in web-mediated interaction and group learning: Theoretical and empirical analysis. *International Journal of Web-Based Learning and Teaching Technologies, 1*(1), 1-28.

Dabbish, L., & Kraut, R. (2008). Awareness displays and social motivation for coordinating communication. *Information Systems Research, 19*(2), 221-240.

DeGarmo, M. (2004). *Issues concerning integration of unmanned aerial vehicles in civil airspace.* Report sponsored by the Federal Aviation Administration, Project No.: 02044111-06.

DeLuca, D., Gasson, S., & Kock, N. (2006). Adaptations that virtual teams make so that complex tasks can be performed using simple e-collaboration technologies. *International Journal of e-Collaboration, 2*(3), 64-90.

Dixon, S. R., Wickens, C. D., & Chang, D. (2005). Mission control of multiple unmanned aerial vehicles: A workload analysis. *Human Factors, 47*(3), 479-487. doi:10.1518/001872005774860005

Draper, M., Geiselman, E., Lu, L., Roe, M., & Haas, M. (2000). Display concepts supporting crew communications of target location in unmanned air vehicles. *Proceedings of the IEA 2000/ HFES 2000 Congress,* San Diego; *3.85 - 3.88.*

Driskell, J.E., Radtke, P., & Sales, E. (2003). Virtual teams: Effects of technological mediation on team performance. *Group Dynamics: Theory, Research and Practice, 7*, 297-323.

Fern, L., & Shively, J. (2011). Designing airspace displays to support rapid immersion for UAS handoffs. *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting,* 81-85.

Gibson, C. B., & Gibbs, J. L. (2006). Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation. *Administrative Science Quarterly, 51*, 451-495.

Goyer, R. (2010, May 12). Safety against the odds. *Flying Magazine*. Retrieved February 3, 2012 from http://www.flyingmag.com/safety/training/safety-against-odds.

Hambley, L. A., O'Neill, T. A., & Kline, T. J. (2007). Virtual team leadership: The effects of leadership style and communication medium on team interaction styles and outcomes. *Organizational Behavior and Human Decision Processes, 103*(1), 1-20.

Hing, J., Oh, P., & Hing, J. (2009). Development of a UAV piloting system with integrated motion cueing for training and pilot evaluation. *International Symposium on Unmanned Aerial Vehicles 2008, 3-19.*

Kalinowski, N. (2010, July 15). Testimony before the House of Representatives, Washington, D.C. Retrieved on March 1, 2012 from http://www.faa.gov/news/testimony/news_story.cfm?newsId=11599.

Kiekel, P.A., Gorman, J.C., & Cooke, N.J. (2004). Measuring speech flow of co-located and distributed command and control teams during a communication channel glitch. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, 683-687.

Lacher, A., Zeitlin, A., Maroney, D., Markin, K., Ludwig, D., & Boyd, J. (2010). *Airspace integration alternatives for unmanned aircraft*. AUVSI's Unmanned Systems Asia-Pacific 2010 Singapore, 2010.

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.

Marks, M.A, Mathieu, J., & Zaccaro, S. (2001). A temporally based framework and taxonomy of team processes. *The Academy of Management Review, 26*(3), 356-376.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*, 273-283.

Maynard, M. T., Mathieu, J. E., Rapp, T. L., & Gilson, L. L. (2012). Something(s) old and something(s) new: Modeling drivers of global virtual team effectiveness. *Journal of Organizational Behavior, 33*(3), 342-365. doi: 10.1002/job.1772

Mouloua, M., Gilson, R., Kring, J., & Hancock, P. (2001). Workload, situation awareness, and teaming issues for UAV/UCAV operations. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting, 162-165.*

Neville, K., Blickensderfer, B., Luxion, S., Kaste, K., & Archer, J. (2012). A cognitive analysis of the UAS human-machine interface and applicable FAA Airworthiness Regulations and guidance. *AUVSI's Unmanned Systems North America 2012 Conference Proceedings*.

Nullmeyer, R.T. (2009). *Predator mishap trends and interpretations by experts: Implications for Implications for training*. Paper presented at the TAAC 2009, New Mexico State University.

Office of the Under Secretary of Defense (2004) .*Uninhibited Combat Aerial Vehicles.* Washington, D.C. 2031-3140.

Powell, A., Piccoli, G., & Ives, B. (2004). Virtual teams: A review of current literature and directions for future research. *The DATA BASE for Advances in Information Systems 35*, 6-39.

Rash, C., LeDuc, P., & Manning, S. (2006). Human factors in US military unmanned aerial vehicle accidents. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research: Vol. 7*,(pp. 117-131). Emerald Group Publishing Limited.

Reynolds, R.T., & Brannick, M. (2009). Effect of communication media on developmental relationships: Self-reported and observed behaviors. *Computers in Human Behavior*, *25*, 233-243.

Rusman, E., Bruggen, J., Cörvers, R., Sloep, P., & Koper, R. (2009). From pattern to practice: Evaluation of a design pattern fostering trust in virtual teams. *Computers in Human Behavior, 25*(5), 1010-1019.

Salas, E., Cooke, N., & Rosen, M. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*, 540-547.

Salas, E., Nichols, D., & Driskell, J. (2007). *T*esting three team training strategies in intact teams: A meta-analysis. *Small Group Research, 38*, 471-488.

Schiller, S. Z., & Mandviwalla, M. (2007). Virtual team research: An analysis of theory use and a framework for theory appropriation. *Small Group Research, 38*(1), 12-59.

Sharma, S. & Chakravarti, D. (2005).UAV operations: An analysis of incidents and accidents with human factors and crew resource management perspective. *International Journal of Aerospace Medicine, 49(1),* 29-36.

Staples, D. S., & Webster, J. (2008). Exploring the effects of trust, task interdependence and virtualness on knowledge sharing in teams. *Information Systems Journal, 18*(6), 617-640.

Stone, N. J., & Posey, M. (2008). Understanding coordination in computer-mediated versus face-to-face groups. *Computers in Human Behavior, 24*(3), 827-851.

Strang, A. J., Knott, B. A., Funke, G. J., Russell, S. M., Miller, B. T., Dukes, A. W., . . . Bolia, R. S. (2011). Collaboration technologies improve performance and communication in air battle management. *Military Psychology, 23*(4), 390-409.

Timmerman, C. E., & Scott, C. R. (2006). Virtually working: Communicative and structural predictors of media use and key outcomes in virtual work teams. *Communication Monographs, 73*(1), 108-136.

Van Breda, L., Jansen, C., & Veltman, H. (2005). Supervising UAVs: Improving operator performance by optimizing the human factor. In D. Schmorrow (Ed.), *Foundations of Augmented Cognition: Vol 11* (pp. 1190-1198).

Van Erp, J. B.(2000). *Controlling Unmanned Vehicles: the Human Factors Solution*. Ft. Belvoir: Defense Technical Information Center.

Walther, J. B., & Bunz, U. (2005). The rules of virtual groups: Trust, liking, and performance in computer-mediated communication. *Journal of Communication, 55*(4), 828-846.

Williams, K. W. (2006). Human factors implications of unmanned aircraft accidents: Flight-control problems. In N. Cooke, H. Pringle, H. Pedersen, & O. Connor (Eds*.). Human factors implications of unmanned aircraft accidents: Flight-control problems* (pp 105-116). Amsterdam, Netherlands:Elsevier.

Williams, K.W. (2004). *A summary of unmanned aircraft accident/incident data: Human factors implications*. U.S. Department of Transportation, Federal Aviation Administration, Office of Aviation Medicine, Washington, DC.

Xiang, H., & Tian, L. (2011). Development of a low-cost agricultural remote sensing system based on an autonomous unmanned aerial vehicle (UAV). *Biosystems Engineering, 108*(2),174-190.

NORTH-UP, TRACK-UP, AND CAMERA-UP NAVIGATION OF UNMANNED AIRCRAFT SYSTEMS

Sheryl L. Chappell
National Transportation Safety Board
Washington, DC
R. Brian Valimont
Human Factors Center of Excellence, Honeywell Aerospace
Phoenix, AZ

To optimize UAV reconnaissance operations, direction of viewing and direction of travel must be allowed to diverge. Our challenge was to design a control and display strategy to allow the operator to easily look where they're going, go where they're looking, and look and go in different directions. Two methods of control were devised to align traveling forward, viewing forward and commanding forward. The operator can command the UAS to turn to camera or command the camera to point in line with the direction of travel (eyes forward). We have also introduced a new camera-up map orientation. The operator can easily cycle through North-up, track-up, and camera-up to provide the best link between the exo-centric and ego-centric frames of reference. Ego-centric and exo-centric perspectives allow the operator to combine or separate the vehicle's movement and the camera's view to optimize the search task while maintaining situation awareness of flight hazards.

As humans moving through our environment, we have a natural orientation. We move in the direction of our natural stride and our eyes are oriented in that same direction. We can look from side to side and up and down to insure our movement will not be impeded or unsafe but we always retain our visual orientation as the direction of locomotion. When we operate a manned ground or air vehicle, we retain a natural orientation for movement that is aligned with the orientation of our scan. Our orientation for locomotion is naturally ego-centric.

When operating an unmanned aircraft system (UAS) that can move in any direction, there is no natural orientation. In addition, there are no visual cues or kinesthetic cues such as gravity and momentum to provide an orientation. In most situations, the operator must orient to the direction of movement, using a camera orientated in the same direction to prevent collision while navigating to the desired location. However, in some situations it is necessary to orient the camera in other directions, such as during a search and rescue where the camera is scanning a broad area. In these situations it may be necessary to rapidly change the direction of movement to the direction of the camera, for example when detecting a person to be rescued.

This paper describes a methodology of command and control that enables the operator of a UAS to resolve the conflict between the camera orientation and the vehicle's direction of movement. A combination of ego-centric and exo-centric perspectives allows the operator to combine or separate the orientations of the vehicle's movement and the camera's view to optimize the search task while maintaining situation awareness (SA) of the flight hazards.

Review

Orientation is a key cognitive component when navigating through a three dimensional space. Early studies in cockpit displays have established two frames of reference, ego-centric and exo-centric, from which people draw their orientation (Jensen, 1981; Spradlin, 1987). To accurately maintain orientation, the operator must cognitively couple these two reference frames, which traditionally have corresponded with the map (exo-centric) and the forward view of the world (ego-centric; Aretz, 1991). These basic frames of reference have become established as fundamental and spread from manned aviation, to virtual environments, to teleoperation of UASs (Mintz, Trafton, Marsh, & Perzanowski, 2004).

To navigate, an operator must continually associate local situation awareness with global situation awareness to answer the question, Am I where I should be? The ego-centric frame of reference is generally characterized as a bottom-up, user-centric perspective. It primarily supports an inner loop of control of the axes of aircraft rotation (pitch, roll, and yaw) necessary to stabilize the aircraft (Wickens, Liang, Prevett, & Olmos, 1994). An ego-centric frame of reference also supports short-range navigation tasks (limited to line of sight), such as the maintenance of aircraft heading and obstacle avoidance. This supported information is collected, assimilated, and maintained as localized SA. The exo-centric frame of reference is generally characterized as a top-down, world-based perspective. It supports an operator's need for extended spatial location information. The exo-centric frame of reference supports information characterizing the geographical location of the aircraft, and its relationship to landmarks, terrain, other aircraft, weather, and proposed and alternative flight paths. This supported information is collected, assimilated, and maintained as global SA.

Previous research on glass cockpit flight displays has generally focused more on supporting ego-centric and local SA than exo-centric and global SA (Wickens et al., 1994). However, research and design work have led to the inclusion of a “Track-up” orientation in electronic maps (Aretz, 1991). Track-up has taken the utility of a pure exo-centric frame of reference provided by the North-up map orientation and performed the proper rotation of the map to orient those world-based references on the map to an ego-centric based perspective. The inclusion of track-up has aided pilots in bridging the gap between the pure exo-centric (map) view and the pure ego-centric (out-the-windscreen) view. However, UASs remove the pilot from the vehicle and hence remove the information derived from the out-the-windscreen view. In an attempt to replace this necessary view designers are using cameras to provide the ego-centric perspective.

As UAS design has progressed, it has often departed from a traditional aircraft configuration of nose, wings and tail. Nose and tail have disappeared, and cameras are no longer limited to providing a forward view. With the increased flexibility of vehicle and camera movement comes an unintended consequence. The UAS operator has lost a fundamental and innate characteristic of orientation, front and back. The purpose for most UAS flights is to give the operator a view. UASs fly to see, they do not necessarily need to see to fly. This is a major shift in how the control system should be designed. In a user-centered design approach to UAS operator controls stations, Chappell and Dunlap (2006) determined that the operator needs to be able to “fly the camera,” to best accomplish their search tasks. To provide this capability, we must first determine how the orientation of the vehicle, the direction of movement, and the direction of the camera interface with the operator’s mental model, the navigation displays, and the movement controls.

In the past, the operation of a UAS typically involved two individuals, one to control the vehicle and another to control the camera. The two worked closely together to get the vehicle to the location where the camera could see the area of interest. As vehicles got more automated and the user interfaces made the operation of the vehicle and the camera easier, a single operator was able to control both the vehicle and the camera. When these functions were combined, the necessity to address the difference between the camera orientation and the vehicle’s orientation became very important.

Types of Navigation

The UAS operator needs the option of different methods of navigation: a pre-determined flight route, a point-to-point real-time routing, and directional control also called teleoperation. Flight along a pre-planned route allows the operator to investigate terrain and obstacle hazards prior to flight and plan for wind and visibility impacts. Specifying a point-to-point routing in real time is more flexible than following a flight plan and less workload than teleoperation. This method of navigation involves commanding a location and may best be performed by selecting a waypoint on an electronic map. For a discussion of this type of navigation using an interactive point-and-click approach see Chappell, 2007. Teleoperation involves commanding a direction in four dimensions by specifying a continuous speed, heading, and altitude. Teleoperation is most often performed using visual references and requires constant control input for vehicle movement. Neuman and Durlach (2006) have found that a game controller provides better UAS control than a mouse for teleoperation and this type of control has become the recognized standard for modern UASs.

The operator can switch between these three methods of navigation during the flight for optimal effectiveness and efficiency. For example, the routing to and from the area to be surveyed may be pre-planned, a new waypoint may be created to get the UAS near where it is needed based on changes in the situation. During the actual observation of the area or object of interest, the operator may choose to use manual control which provides the most flexibility in vehicle location.

Types of Maps and Tasks

Research has shown that the type of map can influence the performance of different tasks. For example, Aretz (1991) found that a track-up alignment is better for determining which way to turn because it eliminated the need for mental rotation. With a north-up alignment participants recalled the location of landmarks more accurately. Maps can also be either two-dimensional or three-dimensional. Two-dimensional maps provide an accurate representation of distances and therefore are best for navigation using a route plan. Three-dimensional perspective maps more closely align with the image from the video and therefore are best for determining the location of objects in the field of view. The video and the map can be combined on the operator’s display. Calhoun, Ruff, Lefebvre, Draper and Ayala (2007) found that a picture-in-picture format where the video was overlaid on a synthetic terrain map reduced the search time for landmarks. Drury, Richer, Rackliffe, and Goodrich (2006) found that search performance was superior with an orthorectified image overlaid on a map. An operator interface that provides all types of maps and methods of combining them with the video is the approach that the authors recommend to provide the best interface for the three methods of navigation and the visual search tasks.

Orientation Conflicts

The biggest challenge for a single operator of both the vehicle and the camera is the combined task of teleoperation and visual search. The operator needs to be able to explore an area, moving around in it, while maintaining awareness of the vehicle's location and clearance from obstacles and restricted airspace. The user interface designer must determine how the orientation of the vehicle, the direction of movement, and the direction of sensing interfaces with the operator's mental model, the navigation displays, and the movement controls. If the camera orientation is moved away from the vehicle direction of movement, there are three options to proceed with manual direction control: 1) automatically converge the direction of movement with the direction of the camera such that commanding a forward motion turns the vehicle to the direction of the camera, 2) keep the direction of movement and the camera direction independent, and 3) cause the convergence of the direction of movement and the camera direction to be based on operator input.

Video games require the operator to control vehicles or personnel locomotion and allow both ego-centric and exo-centric camera views in all directions. Our research on these games revealed two approaches to resolving the differences between camera orientation and vehicle orientation. The most common approach was to permit the person/vehicle to move independently of the camera direction. Some games, however, converged the two orientations such that as the camera was panned, the vehicle would turn to follow; commanding forward motion with the joystick caused the vehicle to move in the direction of the camera.

A UAS in service in combat (Carey, 2007) allows the operator to continue in the direction of movement after panning the camera away from forward; however the first input to the movement control results in a change in direction to that of the camera, at a fixed speed. For example, the operator may be travelling north at ten miles per hour, turn the camera to the west, and any input to the movement joystick results in an immediate turn to the west at a commanded speed of two miles per hour. This vehicle can also be flown by selecting a point within the video field of view and commanding the camera and vehicle to that direction.

These approaches provide an effective but inflexible means of converging the direction of travel with the direction of viewing. If the UAS operator is always going in the direction of the camera the control system design would be straight-forward, however this poses an undesirable and unnecessary restriction on the operation. Not only do UAS operators need to look around as they move in a particular direction, they also need to look in a particular direction while travelling in a different direction. A simple example is a search pattern along a road; the vehicle travels in the direction of the road while the camera is pointed down or to the side.

New Contribution

Our challenge was to provide an interface that preserves both the camera and vehicle orientations with a mechanism to synchronize the two perspectives using the map and the video. We developed a display philosophy that integrated the orientation of both the camera and the direction of travel in the *video* and the *map*. We also developed a control approach to not only integrate the two directions but also permit divergence with ease.

Figure 1. Image from camera video with direction of vehicle track arrow and compass arrow.

Figure 1 shows a typical image displayed to a UAS operator from the camera video. Figure 2 shows the typical map in a track-up orientation as would be found in current UAS displays, as well as in manned aviation glass cockpits. The triangles show the field of view. In addition to this track-up map view, our user interface allows the operator to easily switch to a new map option, a camera-up orientation to align the map with the "out-the-window" view provided by the camera. This new map orientation eliminates the need for the operator to mentally rotate the map to align with the camera view. Figure 3 shows the map in this camera-up orientation. Even in this static representation it is evident that the landmarks in the video field of view (Figure 1) are more easily matched with those on the map, thereby achieving integration between ego-centric and exo-centric perspectives. Note that the video in Figure 1 contains an overlay of the direction of travel (the black arrow in the lower center) and the compass arrow indicating north. The maps contain the arrow for the direction of travel, the triangle for the field of view of the camera, and the north arrow. These cues further help to integrate the two perspectives within the framework of each individual perspective.

Figure 2. Track-up map display as the operator turns to fly in the direction of the sensor (teal triangle).

Figure 3. Camera-up map display as the operator turns to fly in the direction of the sensor (teal triangle).

During teleoperation while performing a visual search, the operator is actively diverging the direction of travel and the camera view. The camera now provides the synthetic ego-centric perspective which is not necessarily forward along the direction of travel. This capability poses a significant potential for disorientation. The expectation is that commanding a forward direction on the control matches travel in the direction of the camera view. When the results of the command inputs do not meet the operator's expectations, a quick way to synchronize the two is required. To accomplish this, we have created a control which we labeled "turn to camera." When the operator selects this option, the direction of travel aligns with the camera's aim point. See Figure 2 for the sequence as the track-up map turns with the vehicle. See Figure 3 for the sequence as the camera-up map symbology turns with the vehicle. The controls align with the new direction of travel. Viewing forward, travelling forward, and commanding forward are all aligned.

This control feature gives the operator a method to command the UAS to "go where I'm looking." It is also important to command the camera to "look where I'm going" but then to quickly return to the previous aim point, "look back." The interface we have designed has an eyes-forward function with a complementary command to return to the previous view.

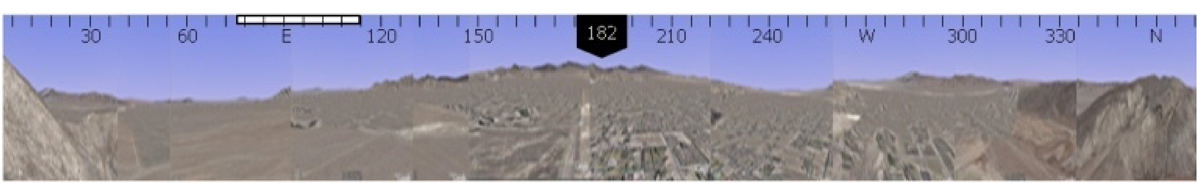


Figure 4. Panorama display. Compass tape is centered on vehicle direction of travel. White band represents camera field of view.

In addition to the camera's field of view for searching, we have designed a panorama display (Figure 4) that is centered on the direction of travel. (See Chappell, 2007 for further description.) The compass overlay also depicts the field of view and direction of the camera (white band). The panorama is accomplished by periodically taking a 360 degree sweep at a horizontal angle. The image is split such that the direction of travel is in the center and the opposite direction is shown at the right and left edges. The panorama image not only provides orientation, but is important for collision avoidance, especially in environments such as urban canyons.

Discussion

Research has shown that the task dictates the best map orientation: North-up, track-up, two-dimensional, and three-dimensional. Our design allows an easy transition between these map types to optimize the operator's task performance.

Our review of the research on vehicle control and orientation combined with our investigation of the current UAS interfaces has illuminated the flaw in the integration between map and video views. This is especially acute when the operator is actively commanding the vehicle movement and camera aimpoint. Our goal was to give the operator maximum flexibility in the vehicle's movement and the view the camera provides and a method to link the situation awareness that the operator derives from both views into one comprehensive mental model. To provide this capability we have a new map orientation labeled camera-up and have included directional cues in the map and video to link the two. We also recognized the need for new controls such as "turn to camera" which realign the vehicle directional controls with the camera aimpoint. Initial simulation trials have shown that the addition of a camera-up map orientation and the ability to take a positive action to realign the manual controls with the direction of the camera constitute a significant contribution to the control of UASs. This combination of ego-centric and exo-centric perspectives allows the operator to combine or separate the orientations of the vehicle's movement and the camera's view to optimize the search task while maintaining situation awareness of the flight hazards.

References

Aretz, A. J., (1991). The design of electronic map displays. *Human Factors, 33(1),* 85-101.

Calhoun, G. L., Ruff, H., Lefebvre, A, Draper, M.H., Ayala, A (2007) "Picture-in-Picture" augmentation of UAV workstation video display. In *Proceedings of the Human Factors and Ergonomics Society*, 70-74.

Carey, B. (2007). *Miniature Air Vehicles*. Retrieved on February 14, 2010 from http://avtoday.com/av/categories/bga/8779.html.

Chappell, S. L. (2007). UAV Control with Pictures and Pointing *Proceedings of the American Institute of Aeronautics and Astronautics Conference and Exhibit.* Reston, VA: American Institute of Aeronautics and Astronautics.

Chappell, S. L. & Dunlap, K. L. (2006). Incorporating operator situation awareness into the design process: A key ingredient for future combat unmanned rotorcraft. *Proceedings of the American Helicopter Society 62nd Annual Forum*. Alexandria, VA: American Helicopter Society International.

Drury, J. L., Richer, J., Rackliffe, N., and Goodrich, M. A. (2006) Comparing Situation Awareness for Two Unmanned Aerial Vehicle Human Interface Approaches. In *Proceedings of the SSRR 2006 Conference*, National Institute of Science and Technology (NIST).

Jensen, R.S. (1981). Prediction and quickening in perspective displays for curved landing approaches. *Human Factors, 23,* 333-364.

Mintz, F.E., Trafton, J.G., Marsh, E., & Perzanowski, D. (2004). Choosing frames of reference: Perspective-taking in a 2D and 3D navigational task. In *Proceedings of the Human Factors and Ergonomics Society,* 1933-1937

Neumann, J. and Durlach, P. J. (2006). Effects of Interface Design and Input Control Measures on Unmanned Aerial System Operator Performance. *Interservice/Industry Training, Simulation and Education Conference (I/ITEC)*, Paper 2882.

Spradlin, R.E. (1987). Modern air transport flight deck design. *Displays,* 171-182.

Wickens, C.D., Liang, C.C., Prevett, T., & Olmos, O. (1994). Egocentric and exocentric displays for terminal area navigation. In *Proceedings of the Human Factors and Ergonomics Society,* 16-20.

EXAMINING MEMORY FOR SEARCH USING A SIMULATED AERIAL SEARCH AND RESCUE TASK

Brandon S. Perelman
Michigan Technological University, Department of Cognitive and Learning Sciences
Houghton, MI
Shane T. Mueller
Michigan Technological University, Department of Cognitive and Learning Sciences
Houghton, MI

In this paper, we report on the development of a synthetic task environment (STE) representing wilderness search and rescue using unmanned aerial vehicles (UAVs) for investigating human unmanned aerial search behavior. Participants navigated using a north up topographical map and detected targets using a more detailed track up satellite image representing the view through the UAV's camera. Participants then completed (1) a path reconstruction task and (2) a memory test in which they indicated locations where they found targets. These tasks aim to address two information types that map onto distinct visual processing pathways afferent to the hippocampus. We discuss example applications using this paradigm, including several methods for scoring memory and navigation performance. Finally, we discuss how the STE enables assessment of the effects of combining or separating pilot and sensor operator roles, search behaviors and strategies, and other human factors limitations faced by operators in aerial search tasks.

Wilderness search and rescue (WiSAR) is the search for missing persons in remote environments. WiSAR operations are often carried out by volunteers on foot aided by resources such as canines, rotary and fixed wing aerial platforms, and marine assets. WiSAR operations begin with the report of a missing person. WiSAR personnel begin by establishing the point last seen, then create a map of probability for target locations distributed according to factors including terrain features, characteristics of the missing person, and current weather (Ferguson, 2008; Lin & Goodrich, 2010; Perkins, Roberts, & Feeney, 2003). Personnel and other assets then begin to search those locations and update the probability map, typically in a Bayesian fashion that considers both the background probability imposed by the terrain and the case based evidence about the missing person (e.g., Lin & Goodrich, 2010).

Recent research has focused on the use of unmanned aerial vehicles (UAVs) in WiSAR (Adams et al., 2007; Adams et al., 2009; Goodrich et al., 2008; Goodrich, Morse, Engh, Cooper, & Adams, 2009). While manned aerial search in general is beneficial to WiSAR operations, the costs of manned aircraft often exceed the resources of WiSAR organizations, which are staffed largely by volunteers (Hoekstra, M. [West Michigan Search and Rescue], personal communication, July 25, 2012). Therefore, WiSAR is an ideal domain for unmanned aerial search.

A typical UAV team in WiSAR consists of a navigator, who is responsible for directing the UAV, and the sensor or payload operator, who examines the video feed from the UAV's sensor package. UAV teams sometimes also incorporate a third individual who oversees the navigator and sensor operator to increase situational awareness (Adams et al., 2009). These characteristics are also found in military UAV teams; however, WiSAR personnel were selected for the present study as a more convenient and accessible population. The goal in both domains (and perhaps others) is to reduce the personnel required to control a single UAV, with a long term goal that a single operator may eventually control multiple UAVs (for a review of this literature, see Cummings, Nehme, & Crandall, 2006). To this end, Cooper and Goodrich (2008) tested multiple control interfaces and found that, with the appropriate interface changes, one human can effectively control a UAV.

One platform for studying unmanned aerial search is the open source Aviones UAV flight simulator (http://aviones.sourceforge.net/). The software has been used to simulate both highly automated algorithmic flight (e.g., Bhatia, Graziano, Karaman, Naldi, & Frazzoli, 2008; Collins, Stankevitz, & Liese, 2011) and human controlled flight (Cooper & Goodrich, 2008). We created a lower fidelity synthetic task environment (STE; Perelman & Mueller, 2013; available at https://sites.google.com/a/mtu.edu/aerialste/) in the Psychology Experiment Building Language (PEBL; Mueller, 2013) to approximate the cognitive requirements of unmanned aerial search tasks while affording greater control over experimental variables related to testing. STEs differ from simulations in that the chief goal in their design is not to replicate environmental characteristics, thus while they are lower fidelity than simulations, STEs permit a greater deal of experimental flexibility and control (Cooke & Shope, 2004). Finally, because the STE is comparatively easy to control and requires very little training to operate, it will permit the inclusion of untrained participants in domain relevant research. One example of this is expert novice differences between untrained undergraduate students and trained UAV pilots. Whereas a high fidelity simulation would require the acquisition and operationalization of novice and expert trained pilots, the present STE would allow untrained participants to fill the control group. While the remainder of this paper pertains specifically to unmanned aerial search in the WiSAR domain, the STE discussed herein may be used to study aerial search in other domains, such as military and law enforcement, as well.

General Features of the STE

The STE consists of two main windows – a north up topographical display with an overlaid probability map, used for navigation, and a track up high resolution satellite image display representing the view through the UAV's sensor package (hereinafter "camera view") (see Figure 1, panel A). At any given time, the entire topographical display is visible, while the camera view display is constricted to only the area directly underneath the UAV. Since the goal of the present study is to analyze search behavior rather than target detection or piloting ability, sensor quality and flight characteristics were approximated arbitrarily to maximize usability during the task. However, these characteristics can be parametrically manipulated to meet specific research requirements.

Gross control of the STE is accomplished either via touchscreen or mouse. Participants navigate the UAV using a carrot and stick method. The UAV orbits the target destination once it arrives until a new destination is selected. Participants designate targets by touch or mouse click inside the camera view display when the target appears, and sound effects provide participants with auditory feedback on hits or misses.

Experimental Tasks

In its current revision, the STE permits two types of tasks: a multiple target search task and a probability density decision test (hereinafter "decision test"). The multiple target search task was created to resemble search and rescue for multiple missing persons using an approximated probability map, while the decision test is intended to test participants' preference for weighing cost versus reward in a more controlled fashion (see Figure 1, panels A and B, respectively). The tasks were parameterized to provide the data necessary for a study in human memory and navigation. However, these tasks may be easily parameterized for additional experiments, and serve as a base from which future experimental tasks may be developed.

During the multiple target search task, participants search for targets, representing lost boy scouts, depicted as blue tents (see Figure 1, panel A). Probable target locations, distributed randomly in each trial, are indicated by blue circles on the north up topographical map. This paradigm is designed to require similar strategic decision heuristics to those required of UAV operators in WiSAR for searching probability maps. Since the present study investigated search behavior and navigation, and not detection, targets were drawn to maximize salience so that if a participant flew over a target, it would be easily spotted. PEBL affords a number of ways with which to adjust target salience in a controlled fashion, such as image alpha (i.e., target transparency), that will permit investigation of probability of detection in future studies. Trial duration is set by a "fuel" variable. In the present study, participants were provided 1,000 fuel units, equating to 79.5 seconds of flight time. This fuel was sufficient for participants to cover roughly 40 % of the total area, assuming that the flight path did not intersect with itself (i.e., areas were not flown over more than once). Fuel was intentionally limited to force participants to make strategic choices during their search.

The multiple target search task contains two subtasks: a path reconstruction task testing spatial memory, in which participants attempt to recreate the path flown by the UAV, and a target memory task testing semantic memory, in which participants indicate which of the possible target locations contained targets (see Figure 1, panels C and D, respectively). These tasks are intended to test performance on tasks requiring processing via dissociated hippocampal afferents, with the path reconstruction and the target memory tasks engaging the dorsal (spatial) and ventral (object) visual processing pathways, respectively (Haxby et al., 1991).

During the decision test, participants search for a single target in two probability regions, one rectangular high density region, representing an oasis, that can be explored in its entirety by flying over a single point, and a long linear low density region, representing a road, for which exploring fully requires a greater time investment (see

Figure 1, panel B). The distances of these two regions from the starting location are parametrically varied (see Figure 2) to test participants' evaluation of temporal cost versus reward.

Data Collection and Analysis

At the beginning of each trial, for both the multiple target search task and the decision test, the STE records target (and foil, in the case of the multiple target search task) locations, the trial number, and parametric and scaling information. Throughout each trial, the STE records the remaining fuel, the UAV's current position, and records the targets that are flown over and reported. Following the multiple target search task, the STE records the coordinates of each point indicated during the path reconstruction task, and a list of the probability locations indicated as containing targets by participants. From these data, the flight trajectory, target information, and the results of the two subtasks can be easily reconstructed.

Data analysis for the path reconstruction subtask of the multiple target search task was accomplished using a path correspondence algorithm (Mueller, Perelman, & Veinott, 2013) that computes congruity between the flown and reconstructed paths (see Figure 3). Since the algorithm simply requires Cartesian coordinates for two paths, it has additional utility as a performance metric in future applications, such as flight formation conformity. Target memory data were analyzed in terms of the temporal serial position in which the targets were discovered.

Demonstration Experiment Methods

Two experiments were conducted to demonstrate using the STE to test participants' performance, and to examine human performance in memory and navigation. Participants in both experiments were drawn from the Michigan Technological University undergraduate participant population. Both experiments were methodologically identical with the exception of the parametric settings during the decision test. In the first experiment (n = 30), the parameters were varied as a 3 x 3 design (road [close / medium / far] x oasis [close / medium / far] in distance from the starting location), whereas in the second experiment these parameters varied according to a 3 x 4 design (road [close / medium / far from starting] x oasis [very close / close / medium / far] in distance from the starting location). The locations of the features during these two experiments are available in Figure 2. In both experiments, participants completed five trials of the multiple target search task in which they searched 12 probability regions, six of which contained targets and six of which were foils (i.e., probability regions not containing targets).

Demonstration Experiment Results

Multiple Target Search Task

Across all trials, participants exhibited a significant improvement in flight performance, as measured both by percent map coverage, F (4, 22) = 4.60, p = .008, and target flyovers, F (4, 26) = 4.32, p = .008 (see Table 1). Generally, participants reported (i.e., clicked) a large percentage of targets that they flew over (M = .87, SD = .22). Across all trials, participants identified a mean of 2.97 targets per trial (SD = 1.32) and remembered a mean of 1.88 (SD = 1.45) of those identified (see Table 1 for specific trial by trial memory scores). Sample results for the path reconstruction task are shown in Figure 3. Since the present study is exploratory and did not use a between groups design, no statistical tests involving the path reconstruction results are presented. However, no statistical relationship was found for performance on the two memory tasks (*results not shown*).

Decision Test

Results of the decision test revealed a strong preference for searching the high density probability region. To test the efficacy of participants' behavior, we developed an optimal model that computes the cost of preferentially searching the high and low probability regions, and then compares these two costs to generate a cost ratio and declare the optimal route. Participants' behavior, in aggregate, correlated strongly with the cost ratio derived from the optimal model in the initial (r = .97, p < .001) and follow up studies (r = .89, p < .001). Participants, in aggregate, probability matched the cost ratio of the two choices generated by the optimal model.

Discussion

The present study describes the development and preliminary evaluation of a STE for unmanned aerial search. Preliminary evaluation indicates that participants generally improve with practice. Additional research is needed to determine consistency with memory effects, such as recency, in the present task and the literature concerning memory in other tasks. Finally, in aggregate, participants appear to be probability matching the cost to reward ratio associated with each route when forced to weigh probability densities.

Potential limitations of the STE include, but are not limited to, the following. First, the STE is not a high fidelity simulation; it attempts to approximate the cognitive requirements of unmanned aerial search in WiSAR. It is possible that this loss of fidelity fundamentally changes the way in which the task is approached cognitively. For example, the STE uses distinct points of probability, whereas probability maps generally produce distributions that are less uniform (i.e., not perfect circular regions). Second, presently, features such as flight characteristics and scale do not map perfectly onto real UAVs. Fortunately, the STE is sufficiently malleable as to permit scaling. Finally, the ecological validity of the STE has yet to be experimentally tested. Therefore, a future study should investigate performance differences between experienced WiSAR professionals and control participants (i.e., college undergraduates).

In future research, the STE described here will allow us to test a number of issues specific to the WiSAR domain and others. Since the STE is highly controlled in its data collection and easy to use, it permits testing of the following phenomena at the basic level with minimal training required. First, the STE will allow us to explore differences between combined and dissociated pilot and sensor operator roles (for an analysis of this problem, see Cooper & Goodrich, 2008). Between groups differences in target identification, memory for the flight path, and memory for targets, can be investigated simply by automating the flight path of the UAV and telling participants that it is being controlled by another user.

Second, the STE allows us to test memory for specific types of targets found during the multiple target search task (i.e., the contents of each high probability region). This is relevant to all aerial search tasks. For example, in actual UAV trials in WiSAR, targets may include clues to the missing person's location rather than the target itself, such as discarded perishables or other items (Goodrich et al., 2009). In military reconnaissance, the precise nature of relevant military targets on the ground may be important for strategic reasons (i.e., remembering the location of an enemy troop transport versus a tank).

Finally, the STE will permit testing the effects of target salience of search behavior and memory. Since the STE can vary the α value (i.e., transparency) of any image object created from a portable network graphics (.png) file, varying target salience is easily achieved in a controlled fashion. Reduced target salience will make signal detection analyses applicable to the data generated by the STE, and probability of detection can be compared between different interface displays and role assignments.

Tables and Figures

Table 1.
Trial by trial data for coverage, flyovers, and memory for targets.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Map Coverage (*M*, *SD*)	.25, .06	.28, .06	.29, .06	.29, .06	.30, .06
Target Flyovers (*M*, *SD*)	2.80, 1.27	3.17, 1.68	3.60, 1.35	3.43, 1.57	3.67, .88
Targets Identified (*M*, *SD*)	2.47, 1.25	2.67, 1.67	3.23, 1.25	3.27, 1.51	3.23, .94
Target Memory (*M*, *SD*)	1.57, 1.28	1.67, 1.73	2.27, 1.48	1.87, 1.61	2.03, 1.16

Table 2.
Number of targets remembered and identified, by serial position.

	1^{st} Target	2^{nd} Target	3^{rd} Target	4^{th} Target	5^{th} Target	6^{th} Target
Identified (*M*, *SD*)	3.70, 1.41	3.07, 1.44	3.00, 1.71	3.11, 1.45	2.44, 1.48	1.81, 1.57
Remembered (*M*, *SD*)	2.22, 1.42	1.74, 1.26	1.74, 1.29	2.00, 1.27	1.74, 1.40	1.15, 1.06
Proportion Remembered (*M*, *SD*)	.59, .34	.55, .34	.56, .41	.64, .36	.69, .46	.44, .47

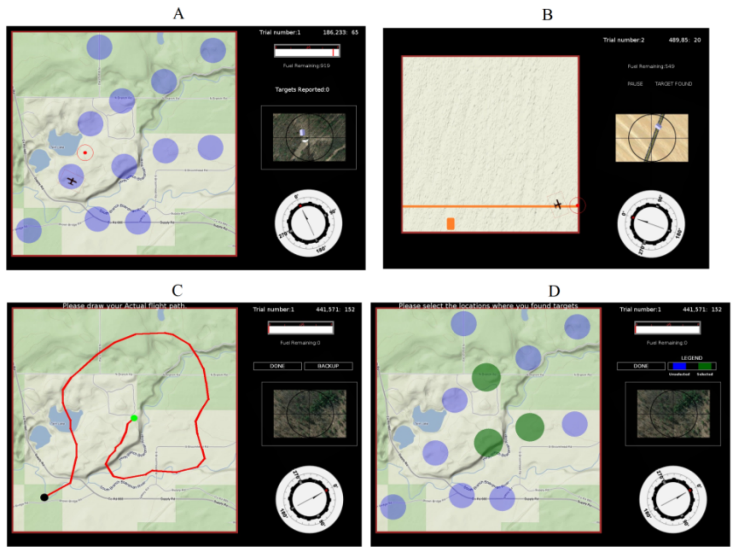


Figure 1. Panel A shows the STE running the multiple target search task. The UAV is depicted with the airplane icon. The UAV's current destination and orbit trajectory are indicated by the red dot and ring, respectively. The north up topographical map (left) permits navigation and depicts possible target locations as represented by the blue circles. Panel B shows the STE running the probability density decision test. The two orange regions represent possible target locations. Panels C and D show the path reconstruction and target memory tasks, respectively.

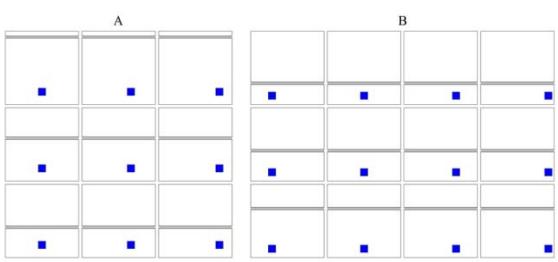


Figure 2. Feature locations for each study. Panel A depicts locations of the road (grey line) and oasis (blue rectangle) in Experiment 1, while Panel B depicts the feature locations in Experiment 2.

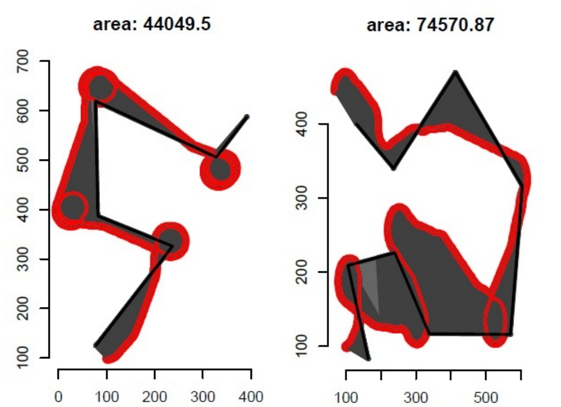


Figure 3. Sample results from the path reconstruction subtask. Participants' flight paths are shown in red, while the reconstructed paths are shown in black. The area of the polygon created by mapping analogous points on each path, shown above each panel, acts as a measure of path congruity and performance

References

Adams, J. A., Cooper, J. L., Goodrich, M. A., Humphrey, C., Quigley, M., Buss, B. G., & Morse, B. S. (2007). Camera-equipped mini UAVs for wilderness search support: Task analysis and lessons from field trials. *BYUHCMI Technical Report, 2007-1*.

Adams, J. A., Humphrey, C. M., Goodrich, M. A., Cooper, J. L., Morse, B. S., Engh, C., & Rasmussen, N. (2009). Cognitive task analysis for developing unmanned aerial vehicle wilderness search support. *Journal of Cognitive Engineering and Decision Making, 3,* 1-26.

Bhatia, A., Graziano, M., Karaman, S., Naldi, R., & Frazzoli, E. (2008). Dubins trajectory tracking using commercial off-the-shelf autopilots. *In Proceedings of American Institute of Aeronautics and Astronautics Guidance, Navigation, and Control Conference and Exhibit*, Honolulu, HI: AIAA.

Collins, G. E., Stankevitz, C., & Liese, J. (2011). Implementation of a sensor guided flight algorithm for target tracking by small UAS. In *Proceedings of SPIE*, Orlando, FL: SPIE. doi:10.1117/12.884314

Cooke, N. J. & Shope, S. M. (2004). Designing a synthetic task environment. In *Scaled Worlds: Development, Validation, and Application*, Schiflett, S. G., Elliott, L. R., Salas, E., & Coovert, M. D. (eds.), Surry, England: Ashgate, 263-278.

Cooper, J. & Goodrich, M. A. (2008). Towards combining UAV and sensor operator roles in UAV-enabled visual search. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, New York, NY: ACM. 351-358. doi:10.1145/1349822.1349868.

Cummings, M. L., Nehme, C. E., & Crandall, J. (2006). Predicting operator capacity for supervisory control of multiple UAVs. In *Innovations in intelligent machines: Studies in computational intelligence*, *70*, Chahl, J. S., Jain, L. C., Mizutani, A., & Sato-Ilic, M. (eds.)

Ferguson, D. (2008). GIS for wilderness search and rescue. In *ESRI Federal User Conference*. Vol. 2012, 10.

Goodrich, M. A., Morse, B. S., Gerhardt, D., Cooper, J. L., Quigley, M., Adams, J. A., et al. (2008). Supporting wilderness search and rescue using a camera-equipped mini UAV. *Journal of Field Robotics, 25,* 89-110.

Goodrich, M. A., Morse, B. S., Engh, C., Cooper, J. L., & Adams, J. A. (2009). Towards using UAVs in wilderness search and rescue: Lessons from field trials. *Interaction Studies, 10*, 455-481.

Haxby, J., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., et al. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. In *Proceedings of the National Academy of Sciences*, *88*, 1621-1625.

Lin, L. & Goodrich, M. A. (2010). A Bayesian approach to modeling lost person behaviors based on terrain features in wilderness search and rescue. *Computational and Mathematical Organization Theory, 16*, 300-323.

Mueller, S. T. (2013). PEBL: The psychology experiment building language (Version 0.13) [Computer experiment programming language]. Retrieved Feb. 2013 from http://pebl.sourceforge.net.

Mueller, S. T., Perelman, B. S., & Veinott, E. S. (2013). An optimization approach for measuring the divergence and correspondence between paths. Manuscript in preparation.

Perelman, B. S. & Mueller, S. T. (2013). A Synthetic Task Environment for experimentally testing human factors effects in unmanned aerial search. Retrieved from https://sites.google.com/a/mtu.edu/aerialste/.

Perkins, D., Roberts, P., & Feeney, G. (2003). *Missing person behavior: An aid to the search manager*. Northcumberland, UK: Centre for Search Research. http://www.searchresearch.org.uk. Accessed October, 2012.

SUPERVISORY CONTROL STATE DIAGRAMS TO DEPICT AUTONOMOUS ACTIVITY

Michael Patzek
Air Force Research Laboratory
Wright-Patterson AFB, OH

Clayton Rothwell
Infoscitex Corporation
Dayton, OH

George Bearden
Air Force Research Laboratory
Wright-Patterson AFB, OH

Benjamin Ausdenmoore
Ball Aerospace & Technologies Corporation
Fairborn, OH

Allen Rowe
Air Force Research Laboratory
Wright-Patterson AFB, OH

As capabilities of autonomous systems expand, traditional geospatial information displays used for supervisory control will likely need to be augmented with more explicit information on higher-order autonomous activities, such as goal-directed task selection, situation assessment, decision-making, and planning. We present a supervisory control interface for higher-order autonomy based on finite state machine diagrams called Layered Pattern Recognizable Interfaces for State Machines (L-PRISM). L-PRISM is a hierarchically arranged set of nested state diagrams coupled with a temporal control and payload viewer. The mission goals and tasks are at the top-most layer and sub-tasks and states at lower layers, providing varying levels of information abstraction. Task diagrams use unique layouts to facilitate pattern recognition, which is anticipated to improve the operator's situation assessment. Furthermore, system behaviors can be viewed in real time or retrospectively evaluated. This paper will describe the L-PRISM concept and plans for examining its effect on supervisory control performance.

The Air Force Research Laboratory's (AFRL's) Value of Information in Collaborative Systems (VICS) initiative is developing technology to enable collaborative control of a flexibly autonomous system. The system is characterized as being composable and fractionated (United States Air Force Chief Scientist, 2010) where heterogeneous members, both human warfighters and unmanned systems, within the system can assemble and combine their respective capabilities and strengths to achieve mission objectives. The heterogeneous components, under varied communication conditions, can coordinate with each other and have the latitude to perform decentralized planning and decision-making as the situation dictates. The goal is to enable agile and adaptive mission management and control for a team comprised of unmanned aerial vehicles (UAVs), unattended ground sensors (UGS), dismounted warfighters with mobile control stations, and an operator located in a central control station. With UAVs equipped and authorized to re-plan and act without human input and under complex, uncertain conditions, a significant technical challenge being addressed is enabling effective supervisory control of the autonomous system. Specifically, the challenge is developing methods for a human operator to sufficiently monitor, inspect, and manipulate the UAVs' activities, which include goal-directed task selection, situation assessment, decision-making, planning, and actions.

Much of the past emphasis in developing multi-UAV supervisory control interface technology has been on effectively portraying the geo-spatial (e.g., a tactical situation map) and vehicle status (e.g., flight parameters, navigation mode, fuel level) information. The associated displays (see example of multi-UAV control station in Figure 1) have transitioned well for many supervisory control concepts as these have typically been applications

Distribution A: Approved for public release; distribution unlimited.
88ABW Cleared 2/15/2013; 88ABW-2013-0861.

where the mission planning and management was done with the operator's direct and near-continuous involvement (Feitshans, Rowe, Davis, Holland, & Berger, 2008; Patzek et al., 2009). However, as future, more autonomous systems assess situations and create their (re-)plans with little or no human involvement, the information typically contained in the tactical situation map and system status displays will not be adequate for the operator to fully understand what the vehicles are doing and why in every case. This, combined with periodic (versus continuous) updates on vehicle locations, status and plans, may lead to significant operator uncertainty and errors in supervisory judgment.

*Figure 1***.** AFRL's Vigilant Spirit Control Station: An Example of a Multi-UAV Control Station. A) Tactical Situation Map, B) Vehicle Summary Information, C) Vehicle Management Panel, and D) Sensor Management.

Given well-known human-automation interaction shortcomings (e.g., mode confusion, vigilance decrement, lack of appropriate feedback, automation complacency, and bias) and foreseeing supervisory control challenges with a dynamic multi-vehicle autonomous system, we are investigating supervisory control interface concepts to display higher-order mission management and control information pertaining to the vehicles' progress towards mission goals, tasks, and the underlying rationale for plan changes during mission execution. The goals, tasks, and rationale are intended to provide a window into the vehicle's perception and assessment of the situation. In addition to providing more detailed information concerning the information processing and behavior of the autonomous system, there is a challenge to present the information in a manner that affords quick and accurate assessment of the multi-vehicle system. With this in mind, our goal is to design concepts that can support the use of symbols and patterns in an attempt to support "at a glance" recognition of complex activities.

We examined design visualization techniques used to illustrate and represent complex, multi-state systems and processes with the idea that perhaps these methods could be made dynamic and animated to represent real-time (and/or past) activities. Diagram and graph methods such as flowcharts, binary decision trees, goal graphs, finite state machine diagrams, and petri nets were assessed for their potential application for an intuitive dynamic display of the autonomous system. Finite state machine diagrams were selected given their efficiency to display multiple systems traversing multiple tasks and states, and the ability to nest states (Harel, 1987) within tasks to represent rules, constraints and overall mission decision logic. Furthermore, the arcs or directional lines between states represent the conditions that must be met by the systems to be able to move from one state into the next, which supports the notion of providing the operator the rationale for autonomous mission, task, and state changes. Another reason for choosing finite state machine diagrams as the basis for the interface was that they can be formed in a manner to produce unique layouts or patterns, supporting the goal to develop an interface that fosters efficient recognition of the activities. Finally, a form of a state diagram user interface has been demonstrated for robot mission planning and representing robot tasks (MacKenzie, Arkin, & Cameron, 1997; Endo, MacKenzie, & Arkin, 2004).

Currently, we are developing hierarchical pattern-oriented state diagram concepts to represent the autonomous activities. In addition, a control timeline and payload viewer are also being developed to navigate and inspect the past events and associated details, including images, videos, audio recordings, and text messages. The layered finite state machine diagrams combined with a control timeline and payload inspection display, collectively referred to as Layered Pattern Recognizable Interfaces for State Machines (L-PRISM), will be integrated with the multi-UAV control station's tactical situation map and system status information to assist the operator to not only be aware of vehicles' locations and planned routes but also their mission goals, associated tasks, and states to achieve the mission goals. The remainder of this paper will describe the L-PRISM concept and plans for its testing and evaluation.

The L-PRISM Concept

The L-PRISM concept, shown in Figure 2, is composed of three display components: A) the State Diagram, B) the Timeline and C) the Payload Viewer. The State Diagram uses the conventions of finite state machine diagrams to represent the autonomous system activity to the operator in real time. The State Diagram depicts the autonomy through the *state nodes* and *transition arcs* of the diagram. State nodes are display elements that represent autonomous tasks, showing what the vehicle is doing. The transition arcs represent the specific criteria for changing tasks, showing why the vehicle may move to a new task. Figure 2 shows the diagram for the "Monitor" task, which has five sub-task or state nodes: Patrol, Isolate, Capture, Deliver, and Response. When a task change occurs, the transition arc is animated with the vehicle color to emphasize the task change and help mitigate change blindness (see Simons, 2000, for a review).

Expanding on the conventions of finite state machine diagrams, multiple vehicles can be displayed in one diagram to accommodate multi-UAV monitoring and control. The State Diagram simultaneously shows the state of each UAV in the mission through vehicle icons. Vehicle icons depict the vehicle type, identify the vehicle by its call sign and unique color, and provide a time-on-task clock. The UAV's call sign and unique color are consistent throughout the rest of the control station and support visual momentum when transitioning between different displays. The time-on-task clock shows how long the vehicle has been working in a specific task. If a vehicle's task changes, its icon moves to a new task. The autonomous system can only make task and sub-task changes that follow the transition arcs. For example, in the Monitor task, Patrol can only lead to Isolate. The operator, however, is able to reassign vehicles to any task or sub-task regardless of transition arcs by dragging and dropping the UAV symbol into the desired task (i.e., direct manipulation). For tasks that have no transition arcs, such as the middle task in Figure 2's State Diagram, task changes can only be made by operators. In general, vehicle control within L-PRISM has the flexibility to support different levels of automation (e.g., management by exception, management by consent, operator directed), as long as there is adequate communication with the particular vehicle(s).

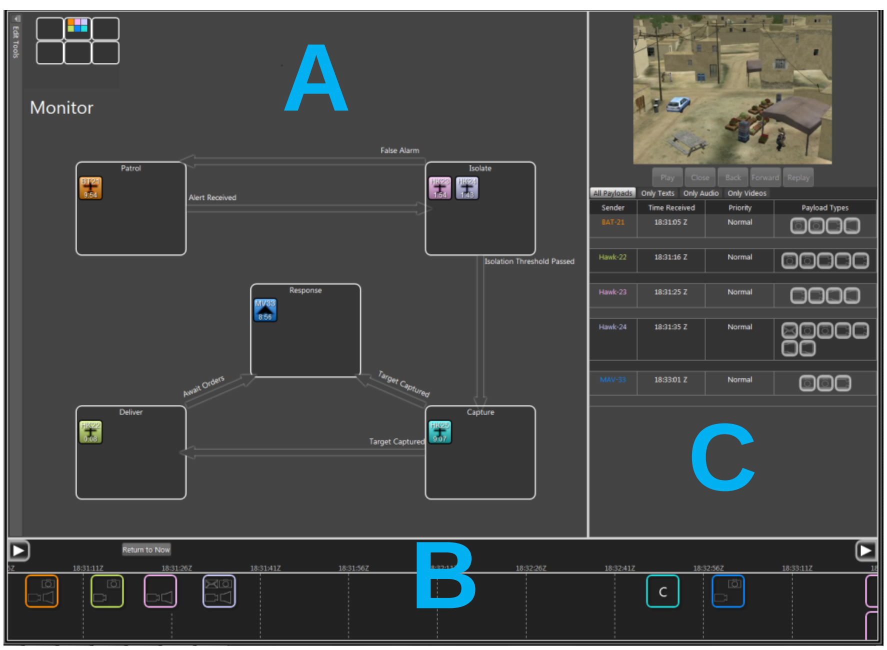


Figure 2. L-PRISM and its three components: A) the State Diagram, B) the Timeline, and C) the Payload Viewer.

Distribution A: Approved for public release; distribution unlimited.
88ABW Cleared 2/15/2013; 88ABW-2013-0861.

Another expansion on the conventions of finite state machine diagrams is in how L-PRISM uses a layered arrangement of nested state diagrams to provide representation of autonomous tasks at varying levels of abstraction. L-PRISM's State Diagram is currently made up of three levels or layers: the Mission Layer, the Task Layer, and the Sub-Task Layer (Figure 3). These levels of abstraction are expected to enhance understanding and management of autonomous activities in part by displaying connections between actions, plans, and goals. The Monitor task diagram (described above and shown in Figure 2) is an example of a Task Layer that shows sub-task components making up the monitoring mission task. At a more abstract level, the State Diagram shows the Mission Layer (see Figure 3) where each mission task (e.g., Monitor, Surveillance, Overwatch) contains its sub-tasks. Active tasks and sub-tasks are identified with UAV icons shown in them. Additional details concerning the tasks and the UAVs performing the tasks can be found by viewing the associated Task Layer. To maintain overall system awareness when accessing a Task or Sub-Task layer, a small higher level state diagram is portrayed in the top left corner of the display. It shows the UAVs' mission task assignments and also functions as method to navigate back to a higher layer. Across all levels, each state machine has a uniquely patterned layout to redundantly code the individual tasks, attempting to facilitate "at a glance" recognition. The Monitor task layer can be identified by its unique "X" layout, which is viewable at abstract and detailed levels.

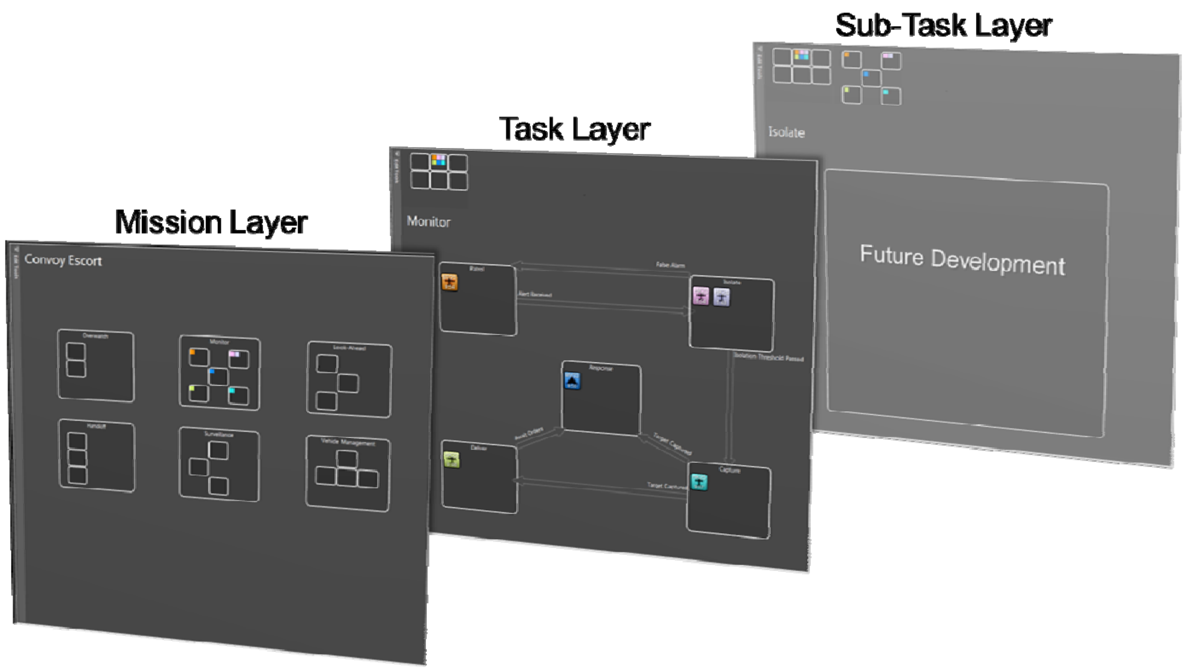


Figure 3. L-PRISM's layers of nested state diagrams showing varying levels of abstraction. The unique patterns of state diagrams are persevered across different layers. At the Task and Sub-Task layer, a small high-level view is presented to maintain some overall system awareness.

Lastly, the State Diagram can be used to build or edit any layer (Mission, Task, or Sub-Task). When in "Edit mode," a tool bar appears that allows new state nodes and new transition arcs to be added to the task diagram. Additionally, all the current state nodes and transition arcs can be moved or deleted. Any changes are snapped to a grid, to facilitate the layout of new patterns. Every node and transition arc has a label that can be edited as well.

The Timeline (see Figure 2) presents information on significant mission events and provides controls to retrospectively review that information across the control station. An event is represented by a colored tile containing a letter or icon that depicts a vehicle task change or payload delivery. Colors match the vehicle color that is used throughout the control station, and icons show the type of payloads. When used as a controller, the time selected on the Timeline updates the State Diagram, the Payload Viewer, tactical situation map and vehicle summary displays with the corresponding information at that time. Operators can proceed forward or backward in time to investigate the tasks and actions of the vehicles and rationale behind those actions using controls for temporal navigation: zoom the time scale in and out, scroll forward and back, step serially through events, and return to current time. This retrospective functionality is in contrast to many timeline applications, which represent what is planned or scheduled to happen. This retrospective was designed to gain better situation awareness of past activities and help in supervisory control of the overall system given challenges associated with periodic communication conditions (e.g., decentralized control). For a vehicle that regained communication with the control station after being out of communication for a period of time, all of its activities will populate the Timeline and are distinguished from other entries by a dashed border. In addition to the retrospective functionality, we plan to expand the timeline to include a view of autonomous system's scheduled tasks and develop a means to contrast the plan with present and past activities. Finally, the Timeline is integrated with the Payload Viewer, such that selection of a payload event on the Timeline will highlight and display that payload in the Payload Viewer.

The Payload Viewer (see Figure 2) displays a sortable list of mission-relevant data referred to as payloads. Typical payloads are images or videos from UAV sensors, but could be operator-generated images or videos as well. Additionally, all operators (i.e., dismounted warfighters with mobile control stations and the stationary central control station) can create and send text messages or "voice note" audio recordings as payloads. Because many of the payloads are an asynchronous form of communication between the system components, the list of payloads was designed to be roughly analogous to email. Each line entry in the list corresponds to a single message that can have multiple payloads attached. The messages and individual payloads show who created the message, what time it was created, the payload type icon, and the message priority. The list of payloads allows sorting and filtering, so that operators can quickly find any payload. When selected, payloads are displayed in the top window with the appropriate controls (e.g., a video payload uses video playback controls).

Future Plans & Summary

The L-PRISM concept has evolved into a functional design and plans are underway for an empirical study to assess the effect L-PRISM has on an operator's supervisory control performance. The plan is to conduct a set of complex, multi-vehicle mission simulations using AFRL's Vigilant Spirit Control Station with and without L-PRISM. The mission context will be reconnaissance, surveillance, and target acquisition related to finding a suspect vehicle within a large road network, referred to as "road monitoring". The UAVs will be operating, for the most part, in a decentralized control manner where they will patrol and isolate the location of the suspect vehicle in collaboration with the unattended ground sensors and dismounted warfighters. L-PRISM will depict the vehicles' tasks and task changes to display the adaptive autonomous systems. The emphasis will be on assessing L-PRISM's impact on the operator's situation awareness and the ability to detect an anomaly or fault (e.g., an incorrect task or action). Mental workload and confidence ratings will also be compared across the conditions.

Two additional independent variables are being considered for the first or perhaps a subsequent study. The first would vary communication between the heterogeneous system members from continuous, real-time communication to periodic. In past evaluations under continuous communication, operators would, on occasion, attempt to infer the automation activities from the tactical situation map. We are interested in examining if the L-PRISM information changes the reliance on the map for assessing the vehicles' tasks and how the communication level affects the operator's behavior. The second variable of interest is the reliability of the autonomous planning and actions and its effect on the use of L-PRISM information given the operator's understanding of the reliability and resulting trust levels. Here the hypothesis is that the L-PRISM display would be relied on more for autonomous systems prone to more situation assessment and planning mistakes as L-PRISM provides a more detailed medium to inspect the higher-order activities.

As part of the VICS initiative, the Vigilant Spirit Control Station and L-PRISM design will be integrated with decentralized control algorithms for a multi-UAV flight test to demonstrate the road monitoring mission task. The plan is to place unattended ground sensors along a large road network and have the UAVs collect information from them and locate vehicles detected by the UGSs or dismounted warfighters. The L-PRISM information will be available to both the central control station and the mobile control stations used by the dismounted warfighters. L-PRISM is expected to provide critical information to understand the complex conditions and activities associated with this composable and fractionated autonomous system.

In summary, L-PRISM is an evolving supervisory control display concept to enable an operator to quickly gain situation awareness and effective oversight of adaptive autonomy and the associated systems: their mission goals, tasks, states, and the underlying rationale for adaptive plans and actions. L-PRISM shows promise for providing many of the desired attributes and features for displaying information on a highly autonomous multi-vehicle system. It supports scaling of the team size, the heterogeneous systems and capabilities, assorted mission tasks and embedded states, both real-time and periodic communication conditions, the patterns for recognizing the tasks, states and overall situation, and can and will be integrated within an existing multi-vehicle supervisory control station.

References

Endo, Y., MacKenzie, D., & Arkin, R. (2004). Usability evaluation of high level user assistance for robot mission specification. IEEE Transactions on *Systems, Man, and Cybernetics- Part C: Applications and Reviews, 34(2),* 168- 180.

Feitshans, G., Rowe, A., Davis, J., Holland, M., & Berger, L. (2008). Vigilant Spirit Control Station (VSCS) "The Face of COUNTER". *AIAA Guidance, Navigation, and Control Conference and Exhibit*, Honolulu, HI. doi: 10.2514/6.2008-6309

Harel, D., (1987). Statecharts: A visual formalism for complex systems. *Science of Computer Programming,* 8, 231-274.

MacKenzie, D., Arkin, R., & Cameron, J. (1997). Multiagent mission specification and execution. *Autonomous Robots, 4(1),* pp. 29-52.

Patzek, M., Zimmer, D., Feitshans, G., Draper, M., Hughes, T., & Flach, J. (2009). Multi-UAV supervisory control interface technology. In *Proceedings of the 2009 International Symposium on Aviation Psychology*, Dayton, OH.

Simons, D. (2000). Current approaches to change blindness. *Visual Cognition, 7(1/2/3),* 1-15.

United States Air Force Chief Scientist (2010). "Technology Horizons: A Vision for Air Force Science & Technology during 2010-2030." Technical report AF/ST-TR-10-01-PR, AF/ST, Office of the Chief Scientist of the U.S. Air Force. Retrieved from http://www.af.mil/shared/media/document/AFD-101130-062.pdf

ADAPTIVE AUTOMATION FOR MULTIPLE AERIAL VEHICLE SUPERVISORY CONTROL: IMPACT OF CHANGING AUTOMATION LEVELS

Heath Ruff
Ball Aerospace and Technologies Corporation
Fairborn, OH
Gloria Calhoun, Crystal Miller, Chad Breeden
Air Force Research Laboratory, 711th Human Performance Wing
Wright-Patterson Air Force Base, OH

Adaptive automation may help balance system autonomy with human interaction in supervisory control environments. Recent results have demonstrated a benefit of performance-based adaptive automation in a multiple unmanned aerial vehicle simulation. However, these findings may instead reflect an attentional benefit from having the task autonomy level change. A between-subjects experimental design was employed to test this possibility. In one group task autonomy level changed with task performance; in the other group levels changed as a function of time elapsed. The results indicated that performance did not significantly differ between the two groups. However, there were significantly more autonomy level changes in the performance-based adaptive automation group. A follow-on study utilizing a yoked-subject design is recommended.

Advances in automation technology are leading to development of operational concepts in which a single pilot is responsible for multiple unmanned aerial vehicles systems (UASs). In this new vision, the operator's role will be more supervisory, monitoring the highly automated flight of UASs and updating plans/resource allocations as required in response to changing conditions (Scott, Mercier, Cummings, & Wang, 2006). The increased role of automation, however, can have negative impacts such as reduced operator situation awareness, decision biases, and complacency (Endsley & Kaber, 1999; Sheridan & Parasuraman, 2006). The application of adaptive automation where the system flexibly allocates tasks between the operator and the automation may be useful in optimizing the tradeoff of operator involvement and workload. With adaptive automation, when some triggering condition is met (e.g., based on critical event, operator performance, operator physiology, models of operator cognition, or hybrid method), either higher levels of automation (LOAs) are applied to one or more tasks or the number of tasks automated increases; Prinzel, 2003). For example, with a performance-based adaptive automation scheme, as operator task performance degrades under increased workload/cognitive demands, the autonomy level increases as an aid. As performance improves during the trial, the autonomy decreases to a lower level. This keeps the participant more in the loop and minimizes automation induced problems like complacency.

Past research has demonstrated a benefit of performance-based adaptive automation (e.g., Kaber & Riley, 1999; Parasuraman, Cosenzo, & de Visser, 2009). In more recent research, the utility of a LOA adaptation scheme based on each participant's individual performance on multiple task types was demonstrated in a multi-UAS simulation (Calhoun, Ward, & Ruff, 2011b; Calhoun, Ruff, Spriggs, & Murray, 2012). Both the speed and accuracy in completing image analysis tasks were improved when this task's LOA changed across three intermediate levels in response to changes in the participant's performance. This result was in comparison to data recorded during trials in which the LOA remained constant at the lowest level of automation (Calhoun, et al., 2012). This improved task performance with performance-based adaptive automation may, however, also reflect an attentional benefit from having the autonomy level change during trials. With each change, feedback on LOA state was updated in the rightmost console window and a status bar below the map. Also, the configuration of the image analysis task window changed to coincide with the LOA. These changes in the information/appearance of the displays may have had an arousal effect and could have been responsible for past reported improvements in the performance-based adaptive automation condition (Calhoun, et al., 2011b & 2012). The present experiment explored this possibility.

Method

Experimental Design

A mixed-design was utilized. The between-subjects variable was the type of adaptive automation. For one subject group, the autonomy level of an image analysis task was tied directly to the participant's individual performance on multiple tasks. In the second subject group, changes to the task's autonomy level were determined by time into the experimental trial (i.e., not related to task performance). All participants completed three experimental trials with their assigned adaptive-automation condition, as well as three trials with a static automation (image analysis task LOA remained constant during the trial). Each participant's trials were blocked by automation condition (adaptive and static) and the order of the two trial blocks was counterbalanced across participants.

Twenty-four volunteers served as participants (14 males, 10 females; mean age = 28.79 years, SD = 8.73 years). All participants reported having normal hearing, color vision, and vision (or correctable) to 20/20. None were experienced pilots: 10 were employed at a U.S. Air Force Base and 14 were recruited from a paid subject pool (compensated $15/hr). The participants were randomly and evenly assigned to one of the two automation groups.

Multi-UAV Simulation Apparatus

A testbed developed by OR Concepts Applied was employed as it facilitates experimental manipulation of task LOA (ORCA; Johnson, Leen, & Goldberg, 2007). Also, this Adaptive Levels of Automation (ALOA, Version 3.0) testbed incorporates the ORCA commercially available mission planner to provide needed complexity and realism. The simulation's computer was a Dell Precision T7500 Workstation with dual Intel® Xeon® CPU x5550 processors @ 2.67 GHz each, 12.0 GB RAM, and a 1.5 GB PCIe nVidia Quadro FX 4800 graphics card (Microsoft© Windows 7 Ultimate 64-bit Operating System). Two Dell 24 in widescreen monitors provided numerous windows that supported participants' completion of multiple tasks. A keyboard and mouse were used for participant inputs.

Experimental Tasks

Each trial consisted of a series of tasks designed to represent the workload envisioned for multiple autonomous vehicle control. The tasks were also designed such that only a few hours of training were required for naïve participants. More details on each task type and frequency (as well as task order) are available (see Calhoun, Ruff, Draper, & Wright, 2011a). Figure 1 provides an illustration of the formats with labels showing the primary windows utilized for each task. There were approximately 6-7 tasks every minute during each 15-min trial. Some of the task types (change detection, system status, and information retrieval) required monitoring displays and making inputs in response to information displayed. Three other tasks (allocation of imaging tasks to UASs, re-routing UASs, and an image analysis task) employed intermediate LOAs that involved both the operator and automation for completion. For these tasks, the automation was 80% reliable. The LOAs for the allocation and re-routing tasks were constant across trials. In contrast, the image analysis task LOA depended on the automation condition in effect.

Image Analysis Task LOAs and Automation Conditions

Participants were prompted that an image was waiting to be analyzed by the addition of a row in the image task window that included an identifier, time added, vehicle source, and counter showing analysis time remaining. Symbology in a timeline also provided cues of pending images. Participants had 20 s to complete the analysis before the image blanked and the task was recorded as a 'miss.' Task completion began with row selection that called up a photo with 19-26 overlaid green shapes (diamonds, squares, circles, and triangles). Analysis required determining the number of diamonds. The next steps depended on the automation condition in effect, to be described next.

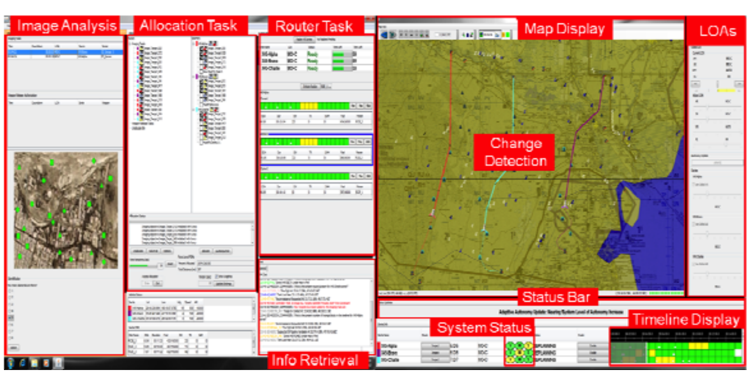


Figure 1. Multiple aerial vehicle supervisory control ALOA testbed showing windows used for tasks.

Static Automation. In this condition, the LOA for the image analysis task was constant throughout trials. A *low* LOA was employed in which the automation presented eight options below the image, each with a different number (see Figure 2). Participants were tasked with selecting the option that corresponded to the number of diamonds in the image (1, 2, …or 8). To complete the task and clear the photo, participants clicked "Select."

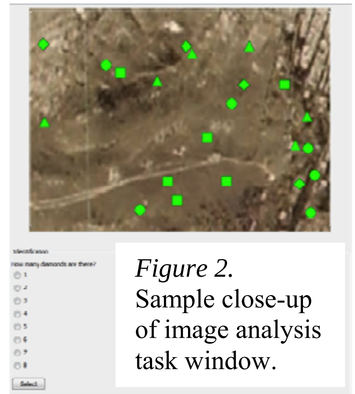


Figure 2. Sample close-up of image analysis task window.

Performance-based Adaptive Automation. Trials started out with the *low* LOA. Each time a criterion task type was completed (allocation of image tasks to UASs, routing of UASs, image analysis, and change detection), its corresponding task completion time was compared to an "expected time window." Instances where task performance was not within thresholds were tallied and once a criterion frequency was met, the LOA changed. (Details on the thresholds are provided in Calhoun, et al., 2012.) As long as the participant's performance remained within threshold, the LOA stayed at the low level. However, if performance exceeded the thresholds, indicating the participant was over-loaded, the LOA increased to the *medium* LOA. At this higher level the automation highlighted its recommended option to assist image analysis and reduce cognitive workload. All options were selectable. If the participant agreed with the automation's recommendation, only the "Select" button needed clicking. However, a different option could be selected. If the participant's performance indicated that the participant was over-loaded with the *medium* LOA (i.e., criterion thresholds exceeded), the LOA increased to *high*. At this level, the automation presented only its recommended option; there was no opportunity to change the selection. To clear the image, the participant's single requirement was to select "Accept" or "Reject." If the image timed-out after 20 s, the system recorded an automatic "Accept." The responses provided a measure of the participant's detection of automation errors, an indication of complacency.

Increasing LOA in response to performance decline was one part of the adaptive cycle. The LOAs also decreased to re-engage the participant if performance indicated that workload had returned to being manageable. Participants were briefed during training that "the system tracks your performance on several tasks to determine if you are over-loaded or under-loaded and raises or lowers the LOA of the image analysis task in response."

Time-based Adaptive Automation. Trials with time-based adaptive automation also started out with the *low* LOA. Throughout the trials, the LOA changed, but the changes were unrelated to the participants' performance. Instead, the changes occurred at experimenter specified times during the trial, and at a similar frequency as that recorded in a previous experiment employing a similar methodology (LOA averaged 3.3 increases and 2.3 decreases across trials; Calhoun, et al., 2012). LOA changes were logical such that, if the LOA was low, a change would make the LOA medium, but it would never immediately transition to high. If the LOA was medium, the LOA might either decrease to low, or increase to high. Participants were briefed on the LOAs and told that the "LOA would change throughout the trial."

Procedures

At the start of the session, each participant completed a background questionnaire and several instruments to measure individual differences (e.g., the 40 Mini-Marker Personality Index to assess the broad traits of the Five Factor Model; Saucier, 2002). A familiarization period followed that took approximately 120 minutes to complete. The testbed's displays and controls were explained, as well as the scenarios and that the vehicles flew automatically along their flight paths. The automation was described as "reliable, but not perfect." Next, each task type was described and practiced in the order of the task's specified priority in a single task environment using the automation condition assigned for the first trial block. This was followed by a series of training trials, gradually increasing the number of task types included in each trial. Training continued until task completion accuracy and response times reached asymptote. Asymptote was defined by mean accuracy and time measures differing by less than 10% on two successive trials that matched the task loading and automation reliability of experimental trials. Next, three 15-min experimental trials with the assigned automation condition were conducted. After each trial, participants completed experimenter developed Likert-type rating scales addressing task difficulty, trust in automation, perceived task performance, situation awareness, workload level, adequacy of automation feedback, and impact of automation on performance. Similar procedures were used for the automation condition assigned for the second trial block. A post-experiment questionnaire was also administered with similar rating scales, as well as questions addressing participants' task completion strategy and comparison of the two automation conditions. The entire session time, including training and questionnaire completion, was approximately 4 hr per participant.

Results

Task Performance Data

The results of an Analysis of Variance (ANOVA) indicated there was not a significant performance difference between the Static (LOA constant at low level) and Adaptive (LOA changed, either performance- or time-based) conditions in terms of task completion time ($F(1,22) = 0.098$, $p = .757$). Results were similar for image task accuracy, but the means showed a larger difference (static: 70.3%, adaptive: 74.2%; $F(1,22) = 3.692$, $p = .068$). There were no other significant main effects or interactions. This included task completion time for the participant group that employed the static and performance-based conditions employed in earlier studies ($p > .731$).

ANOVAs were also conducted for other tasks: allocation of image tasks to UASs, routing UASs, and change detection. Mean time to complete the router tasks was faster with the static condition compared to the (performance- and time-based) adaptive condition ($F(1,22) = 4.326$, $p = .049$). In contrast, mean time to complete the health and status task was longer with the static condition (10.1 s) compared to the adaptive condition (9.4 s; $F(1,22) = 4.607$, $p = .043$). In a comparison of the two participant groups, mean router task accuracy was better with the performance-based condition (93.9%) compared to the time-based one (88.3%; $F(1,22) = 6.707$, $p = .017$).

For the performance-based (PB) adaptive automation condition, the mean number of LOA changes for the 12 participants was 8.0 (mean increases and decreases were 4.7 and 3.3, respectively). These data were significantly higher than that employed in the time-based (TB) adaptive condition. Specifically, this was true for the mean total number of changes (8 PB, 5.7 TB; $F(1,22) = 15.66$, $p = .001$), number of LOA increases (4.7 PB, 3.3 TB; $F(1,22) = 25.919$, $p < .001$), and number of LOA decreases (3.3 PB, 2.3 TB; $F(1,22) = 8.366$, $p = .008$). There was also a statistically significant difference between the performance- and time-based adaptive automation conditions in terms of the mean time spent in each of the three LOAs for the image analysis task (Figure 3; $F(2,44) = 7.071$, $p = .01$). Note that the time spent at each level for the time-based condition was based on experimenter specified adaptive changes. Only the Performance-based Adaptive data were influenced by the participants' performance.

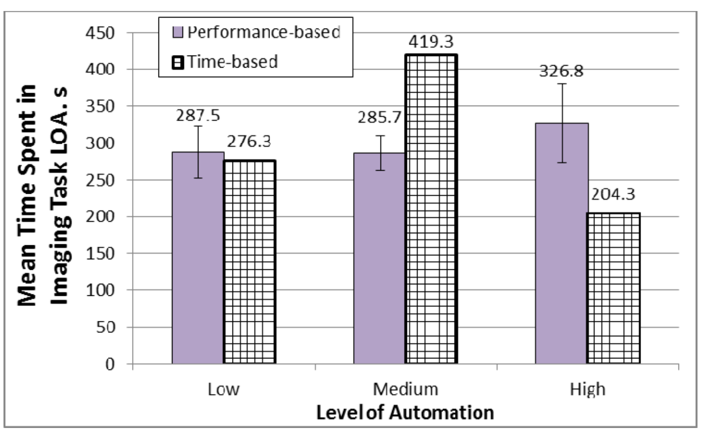


Figure 3. Mean time spent in each level of automation during the image analysis task for the performance-based and time-based adaptive automation conditions. Error bars are standard error of the means in this figure and Figure 4.

Subjective Data

Responses significantly differed for only a few ratings. For a post-trial question on how having the image analysis LOA change affected mental workload, an ANOVA of this between-subjects factor indicated that workload was higher with the Performance-based Adaptive Automation compared to Time-based Adaptive ($F(1,22) = 5.782$, $p = .025$). Responses on a final questionnaire scale comparing Static Automation with the Time-based Adaptive

Automation also differed with 10 of 12 participants indicating that their mental workload was lower with the Time-based Adaptive condition (Kolmogorov-Smirnov non-parametric test: $D(12) = 0.43$, $p < .05$). On this same questionnaire, participants indicated a preference for the high LOA compared to the other LOAs ($D(12) = 0.55$, $p < .01$) and rated the frequency in which the LOA changed to be "Sufficient/About Right" (11 of 12 participants; $D(12) = 0.4$, $p < .05$). This last result is in contrast to a range of responses from the Performance-based Adaptive Automation group: 3 "Slightly Insufficient", 8 "About Right", and 1 "Slightly Excessive" ($p > .10$).

Individual Difference Data

Individual difference measures were subjected to a median-split procedure in which each participant was classified as either a high (above the median) or low (below the median) responder. Each measure was next analyzed with the median-split category (low versus high) as a between-subject factor, along with the automation conditions described earlier. There were two statistically significant results. One pertained to the Emotion factor (emotional stability; neuroticism; $F(1,20) = 6.032$, $p = .023$). Post-hoc tests indicated that participants with high Emotion performed the image analysis task slower with the Performance-based Adaptive Automation compared to the Time-based Automation ($t(10) = 2.585$, $p = .027$; Figure 4). The other result was independent of any automation condition: participants with high Openness were more accurate (77.4%) on the image task analysis compared to participants with low Openness (67.0%; $F(1,20) = 8.651$, $p = .008$). (Openness has been described as a willingness to engage intellectual challenges, and likely to correlate with better performance (Szalma & Taylor, 2011)).

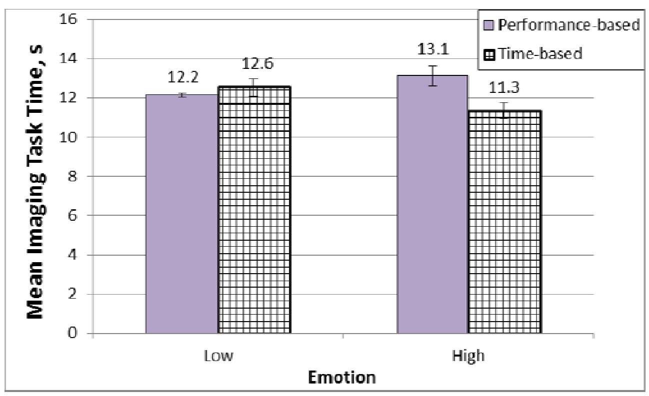


Figure 4. Mean time to complete the image analysis task as a function of automation condition (performance- and time-based adaptive) and participants' Emotion.

Discussion

In earlier research with this simulation, participants' performance on an image analysis task was better with a performance-based adaptive automation control scheme, compared to performance when LOA was static and remained at a low automation level. Thus, it was expected that performance would be better in the present experiment's performance-based adaptive condition, compared to the time-based condition in which the LOA changes were not tied to real-time performance. The results, though, did not support this hypothesis: performance on the image analysis task did not differ significantly between the two adaptive conditions. This finding suggests that any performance benefit from adaptive automation reflects attention benefits from the display updates associated with the LOA changes. However, the additional result showing that the LOA changed significantly more frequently in the performance-based condition compared to the time-based adaptive condition introduces an alternative explanation. It may be that the number of LOA changes also can influence performance. Even if performance-based LOA changes benefit supervisory control in a multi-task environment, an algorithm that is overly sensitive may activate LOA changes at a frequency that, it turn, hampers performance because it is distracting or irritating. Lending support to this supposition is the tendency of participants in the present study with high emotion to perform worse with the condition with the most LOA changes.

Further research is needed to identify the factors contributing to performance benefits of an adaptive LOA scheme. A variety of algorithms should be examined to compare the benefit of LOA changes tied to performance and other trigger alternatives. However, as the present study indicates, the experimental design needs to consider the number of LOA changes as a potentially contributing variable. Use of yoked-subject designs should be considered in which each participant is paired with another on a random basis. For the paradigm used in the present experiment, one participant's frequency of LOA changes (and timing) could feed the scripted changes for the paired participant, without regard to that participant's performance. Experimentation that systematically manipulates the threshold criteria used in a performance-based algorithm for LOA changes would also be informative. With such a detailed analysis, the number and timing of LOA changes both within and across participants could more easily be examined. Subsequent research should also evaluate performance-based adaptive control that involves more than one task in order to explore what combinations of LOAs across several tasks is best, without imposing mode awareness issues.

References

Calhoun, G.L., Ruff, H.A., Draper, M.H., & Wright, E.J. (2011a). Automation level transference effects in simulated multiple unmanned aerial vehicle control. *Journal of Cognitive Engineering & Decision Making, 5*(1), 55-82.

Calhoun, G.L., Ruff, H.A., Spriggs, S., & Murray, C. (2012). Tailored performance-based adaptive levels of automation. *Proceedings of the Human Factors & Ergonomics Society*, 413-417.

Calhoun, G.L., Ward, V.B.R., & Ruff, H.A. (2011b). Performance-based adaptive automation for supervisory control. *Proceedings of the Human Factors & Ergonomics Society*, 2059-2063.

Endsley, M., & Kaber, D. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, *42*(3), 462-492.

Johnson, R., Leen, M., & Goldberg, D. (2007). Testing adaptive levels of automation (ALOA) for UAV supervisory control (Technical Report AFRL-HE-WP-TR-2007-0068), Air Force Research Laboratory.

Kaber, D.B., & Riley, J.M. (1999). Adaptive automation of a dynamic control task based on secondary task workload measurement. *International Journal of Cognitive Ergonomics*, *3*(3), 169-187.

Parasuraman, R., Cosenzo, K.A., & de Visser, E. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology, 21*(2), 270-297.

Prinzel, K. (2003). Team-centered perspective for adaptive automation design (NASA-TM-2003-212154). Hanover, MD.: NASA Center for AeroSpace Information.

Saucier, G. (2002). Orthogonal markers for orthogonal factors: The case of the big five. *Journal of Research in Personality*, 36, 1-31.

Scott, S.D., Mercier, S., Cummings, M.L., & Wang, E. (2006). Assisting interruption recovery in supervisory control of multiple UAVs. *Proceedings of the Human Factors & Ergonomics Society*, 699-703.

Sheridan, T.B., & Parasuraman, R. (2006). Human-automation interaction. In R.S. Nickerson (Ed.), *Reviews of human factors and ergonomics* (Volume 1, Chapter 2, pp. 89-129).

Szalma, J.L., & Taylor, G.S. (2011). Individual differences in response to automation: The five factor model of personality. *Journal of Experimental Psychology: Applied*, *17*(2), 71-96.

A COGNITIVE ENGINEERING APPROACH FOR SHOWING FEASIBILITY MARGINS ON AN IN-FLIGHT PLANNING

Sami Lini[1,2,4]
Bruno Vallespir[1,3]
Sylvain Hourlier[4]
Franck Labat[1,2,4]
Pierre-Alexandre Favier[2,4,5]

[1]Univ. Bordeaux, IMS, UMR 5218, F-33400 Talence, France
[2]CNRS, IMS, UMR 5218, CIH/ISCC team, F-33000 Bordeaux, France
[3]CNRS, IMS, UMR 5218, LAPS, F-33400 Talence, France
[4]HEAL (Thales Avionics/ENSC), F-33000 Bordeaux, France
[5]Univ. Bordeaux, IPB, ENSC, F-33000 Bordeaux, France

The purpose of the ASAP (Anticipation Support for Aeronautical Planning) project was to design an anticipation support for civilian pilots. In this context, we undertook a cognitive engineering approach. Interviews with pilots and in-situ tasks analysis were performed. Helping pilots better anticipate can consist in showing them the room for maneuver for every single task to perform. At first an activity modeling serves as a basis for describing these tasks during a specific phase of the flight (descent/approach). It confirms a need for a visual representation of temporal feasibility margins. According to the constraints of the flight plan (e.g. speed, altitude…) our algorithm dynamically computes the local extreme values for every main flight variables. The tasks performed in this so-defined multivariate tunnel are guaranteed to meet the flight path requirements. The design process and the canvas of our algorithm are presented in this paper. Directions are discussed to evaluate such an algorithm.

The purpose of the ASAP (Anticipation Support for Aeronautical Planning) project was to design an anticipation support for civilian pilots. In this context, we undertook a cognitive engineering approach. Interviews with pilots and in-situ tasks analysis were performed. That led to a model of anticipation which is described as a metacognitive process aiming at providing more temporal and resources margins to perform tasks. On that basis, it is possible to help pilots better anticipate by showing them the room for maneuver for every single task to perform as well as tasks dependencies.

At first an activity modeling serves as a basis for describing the various tasks to be performed during a specific phase of the flight (descent/approach). That model highlights how constrained are these tasks along the main defining flight variables (time, altitude, speed…) and how pilots tackle these constraints. It also confirms a need for a visual representation of feasibility margins (see Lini et al., 2013).

The elaborated tasks graph is used by implementing our algorithm. According to the constraints of the flight plan (e.g. speed, position, altitude…) it dynamically computes the local extreme values for every main flight variables. The tasks performed in this so-defined multivariate tunnel are guaranteed to meet the flight path requirements. A demonstrator of the functioning of this algorithm is implemented. It simulates some of the flight variables computation and shows how the tasks graph is accordingly modified dynamically all along the flight phase.

The design process and the canvas of the algorithm are presented in this paper. Directions are discussed to evaluate such an algorithm.

Activity modeling

For logistic reasons (data availability), we choose to study the descent and approach phases on Rio de Janeiro international airport. At first, an interview is conducted with a flight instructor familiar with this flight. The pilot is required to put into words all the tasks performed in the defined phases. Using the MASK knowledge management methodology (Matta, Ermine, Aubertin, & Trivin, 2002) we conducted a hierarchical task analysis (HTA, Stanton, 2006) aiming at building a knowledge base.

At first, we capitalized the knowledge of an expert pilot on a defined flight phase: the descent and approach phases of a commercial flight are known to be representative of the troubles pilots have to face (Boeing, 2011). This capitalization consists in a recorded semi-directed interview where the subject is asked to describe as accurately as possible his tasks and activity. Questions are asked from time to time to disambiguate grey areas. A camera video recording of the cockpit was also made during an actual flight phase with the expert.

These two sources of information form the material we have been using for building the task diagram. We stressed the inputs and outputs of every single task and their links to anticipation. This allows building a representation of pilots' activity over time. A second model, the activity diagram, gives a higher level of details about it: processes and used resources are detailed. We interviewed a second expert to double check the modeling. Pursuing the goal of designing an anticipation support, both the HTA and pilots-in-the-loop process led to several trends to help better anticipate.

The first one is about helping to compute the feasibility margins (in terms of time and other flight variables) of the most critical tasks to be performed. Two reasonings are considered. The first one is as follows: "given an arrival airport and an ETA, until what time can I consider performing every task in the graph while flying safely and satisfying all the coming constraints". The second one is as follows: "given the current situation, what is the earliest term until which I can consider performing the tasks while still satisfying all the coming constraints".

Another trend is to help dealing with "what-if": to help implementing hypotheses about the flight, in case of diverting for instance. All of this will necessitate a real-time and up-to-date representation of the situation or the ability to dynamically build and represent a projection of an alternative scenario.

Planning algorithm

Objectives

The objective of the presented algorithm is the following: to offer the possibility to visualize and interact with an improved flight plan (predefined but not necessarily specific to a particular flight) which is dynamically synchronized with the current flight situation (position, altitude, speed).

The requirements are as follow:

- Representation of tasks feasibility margins

- To specify a flight plan to a particular airport;
- Automatic spatial and temporal adjustment (ahead/delay regarding an ETA given the current situation)
- What if (simulation and validation)
- Dynamic re-planning in case of diversion

Data structure

The first step of this algorithm consists in modeling a flight plan: elements we want to show and on which computation will be performed. A *flight plan* is an oriented graph in which *flight steps* are vertices while the *change of state* between two flight steps are edges. This plan is built during a two steps process. Firstly it's instantiated from a generic flight plan (built from our HTA previously described) which defines generic procedures for a standard approach. It's then specified through a transformation process which exploits both this generic instance and a constraint list specific to the concerned airport in order to produce a dedicated flight plan specific to the current flight...

Each flight step is defined by:

- a name: it describes the task to be performed;
- the aerodynamic state of the aircraft: it is a mandatory property in order to take into account the plane's shape in the time margins computation. What is the flaps' level? Is the landing gear in or out? Knowing this information will allow adjusting the maximum and minimum plane's evolution speed in the computation of time margins.

Thus the structure of the flight plan describes everything that needs to be performed during the flight and the "shape" of the plane during these very tasks.

A list of constraints can be tied to a flight step. A *constraint* can be either generic (shared between all flight plans) or specific to a flight plan. Managing a flight is not specific to a particular trajectory. Generic rules are used to fly the plane. Like a mental representation, these general rules need to be fulfilled with local and specific information. We then need to distinguish what is generic to any flight and what is specific to a trajectory.

A constraint is defined by:

- a name: it describes the constraint,
- a set of variables (altitude, position, speed, …): depending on the source of this constraint, one or more flight variable can be constrained. The aerodynamic state of the plane will for instance give speed constraints. On the flight path, one might also encounter other speed limitations as well as altitude or position constraints.
- a range of values to reach for each variable: the variable can be constrained to either a single value (pretty scarce) or to a range of values (for instance, speed between 250 kts and 280 kts).

Two kinds of constraints exist:

- a *postrequisite* is a satisfied constraint, the new state of the world, after performing the task, which is necessary in order to define if it is possible or not to move to a next step
- a *prerequisite* is an entry constraint without which the flight step cannot be considered.

Tied to a flight step are a list of its *predecessors* (flight steps) and a list of its *successors* (flight steps). These predecessors and successors allow from every vertex to have a representation of both where it came from and where it goes.

An *arc* is defined as a pair (*p*,*q*) where *p* and *q* are two distinct vertices. A *transition* makes possible to move from a vertex *p* to a vertex *q* if and only if the constraints list of *p* is satisfied and there is an arc between *p* and *q*. Two kinds of arcs are defined:

- a *forward arc* model the progress of the flight plan
- a *backward arc* model the updates of constraints according to the flight plan (aerodynamic state) and the pilot's anticipation.

Given a pair (*p*,*q*), the activation of *p* towards *q* consists in the following steps:

1. *p* is the active flight step
2. Postrequisites are validated: constraints satisfaction is checked
3. If there is an arc between *q* and *p*, *q*'s prerequisites are checked.
4. If *q*'s prerequisites are satisfied, *p* is inactivated and *q* is activated.

Spread algorithms

Two spread algorithms are distinguished:

- The forward spread is used every time the current state of the aircraft is modified. It can also be used in the case of a change of the execution time of a flight step (anticipation).
- The backward spread is always used after a forward spread in order to spread to every flight step the new calculated time limits.

For each vertex, the updated values for each variable are thrown to all of its predecessors/successors (depending on the direction of the spread). The Breadth First Search (BFS) algorithm is used to perform this spread throughout the whole plan.

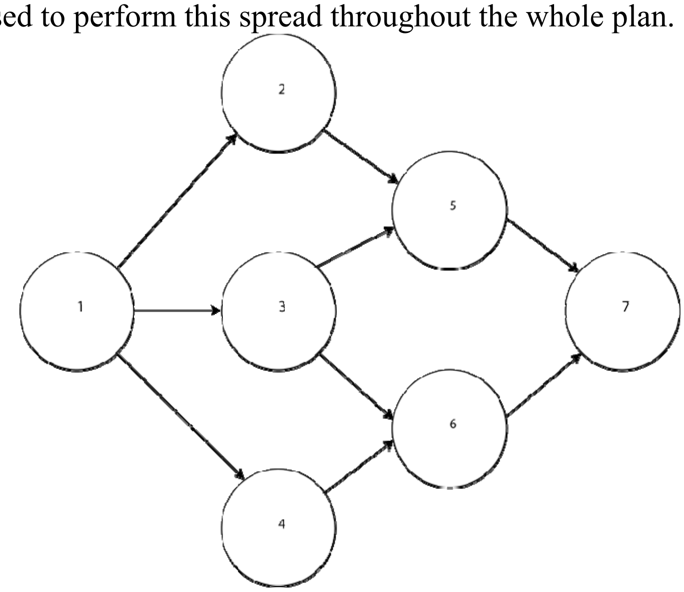


Figure 1 : Steps numbering for a BFS in the increasing order (1 to 7)

The forward spread then consists in:

- A BFS to all forward arcs from an activated vertex

- For each successor, an update operation of all variables using the MaxMin algorithm (see below)

The backward spread consists in:

- A BFS to all backward arcs from an activated vertex
- For each predecessor, an update operation of all variables using the MinMax algorithm (see below)

MaxMin and MinMax algorithms.

The philosophy of the MinMax algorithm is the following: given a vertex of which we know every single characteristic (temporal, spatial, and aerodynamic), we observe each of its predecessors. Every predecessor has one or more constraints.

These constraints are successively and hierarchically analyzed. They are considered in the following order of importance: altitude, speed, vertical speed, position (from HTA). The time to satisfy each of them is calculated using a simplified model of evolution: a constraint defined a state of the world (speed, position…).

We calculate the time it takes to move from this state of the world to the state of its successor. Following the adage "*he who can do more can do less*", we keep the lowest calculated value. The algorithm is then spread to the predecessors using the BFS.

The MaxMin algorithm follows the same philosophy in a forward way.

The actual tasks graph is the following:

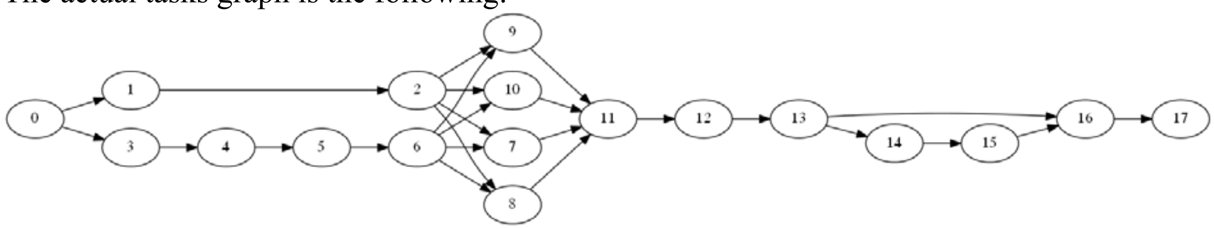


Figure 2: Actual tasks graph

Prototyping

In order to test both the spread equations and the spread models, a prototype HMI was designed. Three panels were defined: a temporal one, presenting all tasks and their time margins, a speed panel, presenting the speed margins in time for each task, and an altitude panel presenting the altitude range for each task.

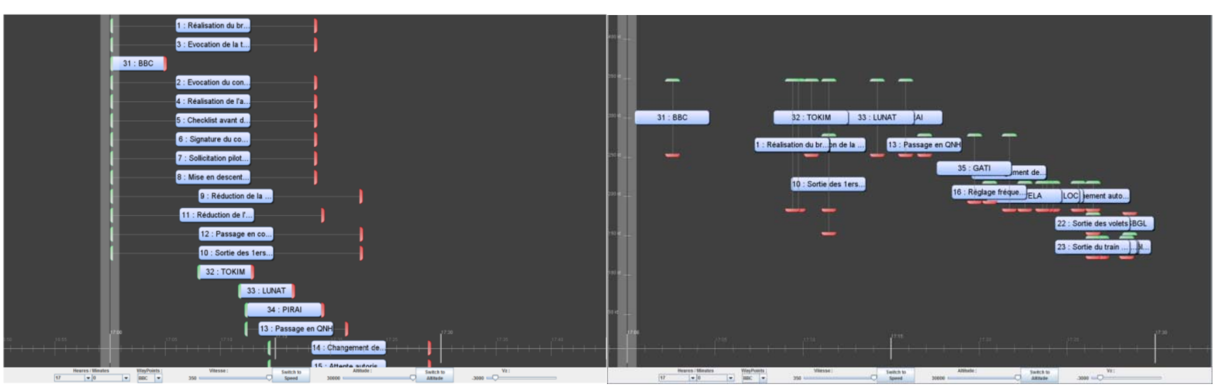


Figure 3: Tasks and speed panels

Conclusion

This algorithm allows displaying a visualization of both the flight path and the tasks to be performed in an inter-linked and up-to-date way. It has been implemented in the ASAP interface which has been evaluated with a panel of 36 commercial pilots in a simulated flight context and showed benefits on both cognitive load and situation awareness (see Lini et al., 2013).

Further improvements of this algorithm shall focus on the plane's energy rather than its independent flight variables.

References

Boeing. (2011). *Statistical summary of commercial jet airplane accidents - Worldwide operations 1959 2011*.

Lini, S., Bey, C., Hourlier, S., Vallespir, B., Johnston, A., & Favier, P.-A. (2013). Evaluating ASAP (Anticipation Support for Aeronautical Planning): a user-centered case study. *Proceedings of the 17th International Symposium on Aviation Psychology*. Dayton, OH.

Matta, N., Ermine, J. L., Aubertin, G., & Trivin, J. Y. (2002). Knowledge Capitalization with a knowledge engineering approach: the MASK method. *Knowledge management and organizational memories*, 17–28.

Stanton, N. A. (2006). Hierarchical task analysis: developments, applications, and extensions. *Applied ergonomics*, *37*(1), 55–79. doi:10.1016/j.apergo.2005.06.003

RISK PERCEPTION IN ECOLOGICAL INFORMATION SYSTEMS

J. Comans, C. Borst, M.M. van Paassen and M. Mulder
Delft University of Technology, Faculty of Aerospace Engineering
Kluyverweg 1, 2629HS Delft, The Netherlands

One issue that regularly occurs in the context of ecological information systems is that these systems can invite operators to migrate to the limits of system performance. This could lead to the assumption that ecological systems are thus inherently unsafe. We argue, however, that the source of this issue is tied to a modeling problem of the work domain. That is, the majority of ecological systems predominantly model the physical or causal structure of the work domain, thereby neglecting the intentional structure. Many complex systems contain a mix of physical and intentional (i.e., rules, procedures, and regulations) that contribute to safe operations of those systems. The work described in this paper examines how visualizing intentional information in an ecological synthetic vision display affects pilot behavior, decision-making, and safety in a terrain avoidance task. An experiment with 16 professional pilots showed that adding an intentional constraint increases the clearance during terrain avoidance and gives them more insight into the terrain avoidance task enabling them to make better decisions.

Cognitive Systems Engineering (CSE) (Rasmussen, Pejtersen, & Goodstein, 1994) and Ecological Interface Design (EID) (Vincente,1999) paradigms are commonly regarded as guiding frameworks to develop ‘transparent’ automation, allowing human agents to monitor the machine and fluently re-direct machine activities warranted by the demands of the situation. The rationale is to let the computer provide the human agent a set of constraints, rather than an explicit solution, that is directly visible on the human-machine interface, within which any action is possible to solve a problem. Such an approach is seen as more robust and resilient, and can be contrasted with a ‘brittle’ approach that provides optimal advice most of the time, but fails spectacularly in a few cases.

Although empirical studies have shown that such information aids enable the human to have a better system understanding and a better notion of the physical limitations, possibilities, and relationships within the work domain, humans often tend to propose actions that are suboptimal, good enough, or even “pushing the envelope” (Rasmussen, 1997). For example, Borst, Mulder, and Van Paassen (2010) showed that their ecological synthetic vision display (SVD) invited pilots to systematically violate minimum terrain clearances. This could lead to the assumption that ecological information systems are unsafe and can ‘promote’ risky behavior.

Although we share the same concern about seeking out the limits of performance, we believe that the risky behavior is tied to the scope of the work domain analysis that is modeled rather than the EID framework itself. That is, the majority of ecological systems predominantly model the physical or causal structure of the work domain, thereby neglecting the intentional structure (e.g., rules, procedures, and regulations) (Hajdukiewicz, Burns, Vicente, & Eggleston, 1999). For example, aviation safety is not only accomplished by the technical systems on board an aircraft, but also by standardized communication and coordination protocols, procedures, and airspace organization. So when the scope mainly includes the causal constraints, the ‘physical structure’ in the environment will be made compelling and this can cause people to pursue these physical boundaries, leaving little room to prevent accidents. On the other hand, however, when the scope is too much on the intentional constraints, the system will generally be safer, but the operational range of physical systems can be significantly limited to effectively solve problems in novel situations. The EID approach, however, can also be used to make both the physical and the intentional constraints visible and it can also manipulate the relative salience of those constraints.

In this paper it is investigated how visualizing intentional constraints in addition to physical constraints affect pilot behavior and decision-making in a terrain avoidance task when utilizing an enhanced SVD. As such, it aims to answer the following question: when pilots are explicitly confronted with intentional constraints in addition to physical constraints, will they make ‘better’ decisions and will they better understand the risks involved in those decisions? The work in this paper is essentially a repetition (in some aspects) of the experiment conducted by Borst, Mulder, and Van Paassen (2010) with the addition of an explicit visualization of the required minimum safe altitude above terrain.

Experiment

An experiment has been conducted to investigate how pilots will respond to adding intentional information to an EID display. Building on the experiment by (Borst et al., 2010), a terrain avoidance task with an SVD was chosen. As an intentional constraint, the minimum terrain clearance was chosen which specifies the minimum height a pilot needs to have above the terrain below the aircraft. The participating pilots were put in situations where climbing over the terrain was the only viable option.

Subjects

A mix of 16 recently graduated and commercial pilots participated. Their average age was 40 years (SD 16.13) with an average experience of 3370 flight hours (SD 3923.07). Four of them were TU Delft/NLR test pilots. One of them was a former military F16 test pilot.

Apparatus

The experiment was conducted in a fixed base flight simulator. The display was shown on an 18 inch monitor located in front of the pilot. An outside visual consisting of fog and cloud fragments was projected on the front and side walls to provide some sense of motion. The aircraft model was controlled by a right hand hydraulic side stick and a throttle quadrant on the left. The throttle contained the trim switch, auto pilot disconnect switch and HSI center button. A mode control panel on top of the instrument panel was used to control the HSI course. A non-linear six degree of freedom Cessna 172 model was used for the experiment. Pitch, roll and throttle commands were directly controlled by the pilot. To compensate for the lack of rudder pedals, a side slip controller was implemented to minimize side slip and engine torque effects. Two different performance settings where used during the experiment. During normal performance runs, the model operated in a normal International Standard Atmosphere giving the normal performance at the altitudes flown. In the reduced performance mode, the aircraft performance corresponded to what would be expected at low density altitude conditions. In this mode, climb performance decreases significantly with altitude.

Display

The display used in the experiment is shown in Figure 1. It is based on a Garmin G1000 NAV III augmented with a synthetic vision system. Three additional cues were added for the baseline display. The flight path vector (FPV) indicating the geometric flight path which provides immediate feedback to the pilot about his current trajectory. If the FPV is pointing at the synthetic terrain, the aircraft will impact the ground at that position if the pilot does not take any further action. The maximum sustained climb angle at full power is shown by a wide green bar (no. 3 in Figure 1) this indication immediately shows whether the aircraft is able to clear the terrain at maximum climb performance. When this line is below the synthetic terrain, the pilot will not be able to climb over it and will have to choose a different maneuver. Similarly the current maximum climb angle (no. 2 in Figure 1) is the maximum climb angle that can be sustained with the current power setting/throttle position. With this indication, the pilot gets immediate feedback about his current climb performance and is able to climb at lower power settings while still being sure to clear the terrain.

The baseline display is augmented with an intentional layer indicating the minimum safe clearance above the terrain (no. 4 in Figure 1). This layer is created by shifting the synthetic terrain up and drawing it in amber behind the physical terrain. In this way, the layer has the same relationship to the FPV as the original terrain. If the FPV is above the layer, the terrain clearance will be at least the required minimum clearance. An additional advantage of adding the intentional layer to the display is that it improves distance perception of terrain features, something that is very difficult in traditional SVDs. Because the layer has a fixed height, its resulting thickness on the display is an indication for the distance to the terrain. Even tough the relationship between distance and thickness is non-linear, it can aide in a crude estimation of the actual distance of terrain features

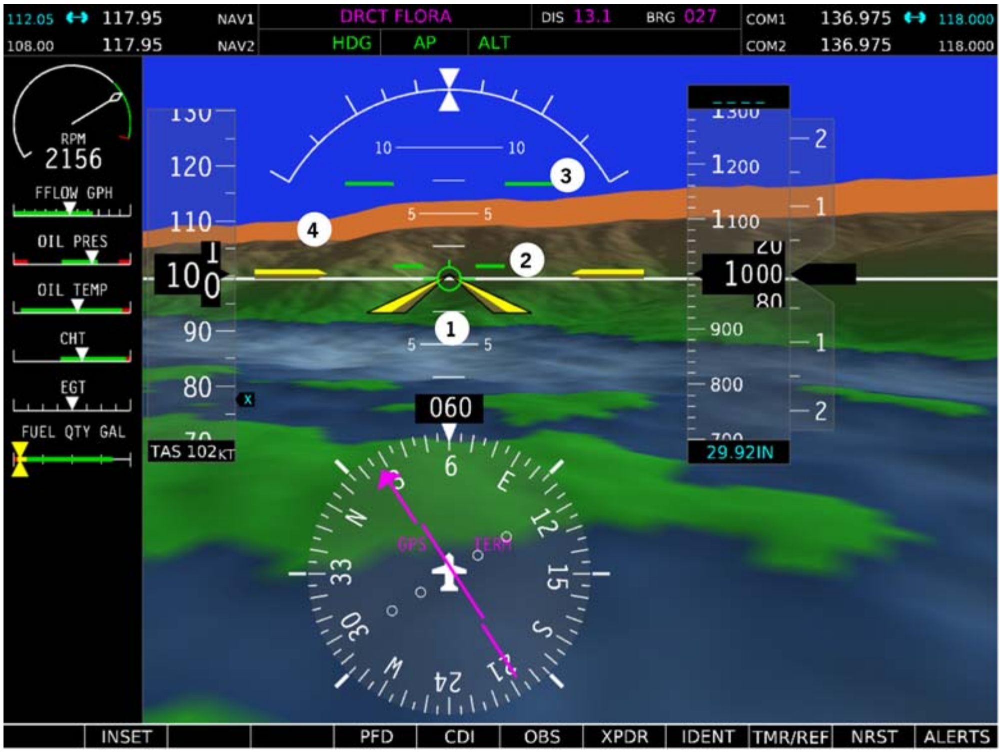


Figure 1: The SVD display, showing: 1) The flight path angle, 2) the current maximum sustained climb angle, 3) the maximum sustained climb angle at full power, and 4) the intentional terrain constraint

Scenario

The scenario used for the experiment consisted of an artificial terrain with a number of narrowing fjords. The base of the fjords was at sea level the tops were around 3000 ft. The pilots were told that they flew into the wrong fjord on their way to the airport and are low on fuel by the time they realize their mistake. Each experiment run started in one of the predetermined initial locations at an altitude below the surrounding fjord tops. From this starting position it was impossible to get to the airport without climbing over the terrain. A navigation beacon was placed in between the initial positions and the airport to provide a navigation reference that could be reached in three to five minutes eliminating the need for long cruise segments to reach the airport.

Pilots were instructed to navigate the waypoint and keep clear of the terrain in a way they considered safe and comfortable. Pilots were not given minimum altitude instructions, but they received a map of the area that showed 4000 ft as the minimum safe altitude for the area they were navigating.

Independent Variables

The experiment used three within-subject variables each having two levels: display configuration, scenario difficulty, and aircraft performance. The display configuration was either a baseline display without the intentional layer, or a display with the intentional layer visible. Two levels of difficulty are used. Easy conditions start with enough margin for a straight climb towards the beacon. The hard conditions required immediate full power and a maximum performance climb to avoid the intentional layer. Two levels of aircraft performance were used. With normal performance, climb performance remains almost constant during the first 4000 ft of the climb. With the reduced performance, climb performance severely deteriorates above 2500 ft.

The order in which the display configuration was tested was used as a between-subject variable. One group of pilots started with 8 runs using the baseline display and then moved the augmented display. The other group started with the augmented display and moved to the baseline display afterward.

Dependent Measure

The main dependent measure of this experiment is the minimum terrain clearance during the experiment run which can be interpreted as a measure of safety of the maneuver.

Next to this objective measure, notes were taken during each run describing the pilots choices. After each run, pilots provided feedback about their strategy and choices.

Experiment Design

The pilots were divided in two groups, one started without the intentional additions, the other started with the intentional additions enabled. Initially there were 18 pilots divided equally among both groups, but during the experiment, two pilots failed to complete the experiment. This resulted in nine pilots starting without intentional additions and seven pilots starting with intentional additions.

Each pilot flew eight conditions per block (two difficulty levels, two performance levels, and two repetitions). The conditions where randomly distributed based on a Latin square matrix to avoid effects based on the condition order. Different Latin squares were used for both blocks. At the end of the experiment, 256 samples were collected (2 * 8 * 16).

Procedure

The experiment started with a training phase to familiarize the pilots with the display, flight controls, and aircraft model. During training, the pilot could fly around freely in a training database and got an explanation on the display features. Once the pilot was familiar with the added features, the measurement runs began. No task specific training was done, only display familiarization.

Before each measurement run started, the pilots were instructed to set the throttle to the trim position. Once the run started, the autopilot maintained altitude and airspeed for five seconds. During this time, the pilot was asked to observe the situation. After the autopilot disconnected, the pilot had to confirm the disconnect by pushing a button and navigate the aircraft towards the navigation beacon. Once they were close to the beacon, the run ended and the pilots provided feedback about their strategy during the run.

At the end of the experiment pilots were asked to complete a questionnaire to evaluate the overall experiment.

Results

To analyze the objective clearance measure, a repeated measures Analysis of Variance (ANOVA) has been performed. To simplify the analysis, the results from the repetitions were averaged per condition resulting in eight data points per pilot. This assumption should not distort the results too much since the majority of the pilots showed a reasonable consistency between the repetitions in terms of strategy and minimum clearance. Figure 2 shows a summary of the minimum clearance values for all pilots per condition together with the mean and standard deviation per condition. The figure shows that the minimum clearance increases when the intentional constraints are visible. This effect is confirmed with the ANOVA showing an effect for the display type ($F(1,14)=5.44$, $p < 0.05$). No significant interactions were found for the difficulty and performance variables with the display type. The display order did not have a significant effect ($F(1,14)=0.687$, $p=0.064$) but was close to the $p=0.05$ significance level. Further research with more pilots might reveal an actual influence of the display order. During the post experiment questionnaire one of the pilots remarked that his strategy without intentional layer had changed by starting with the intentional layer enabled.

The only other significant effects were the main effects of the complexity ($F(1,14)=23.446$, $p<0.01$) and performance ($F(1,14)=15.332$, $p<0.01$) variables. This confirms that the task became more difficult both with decreasing performance and with increasing difficulty.

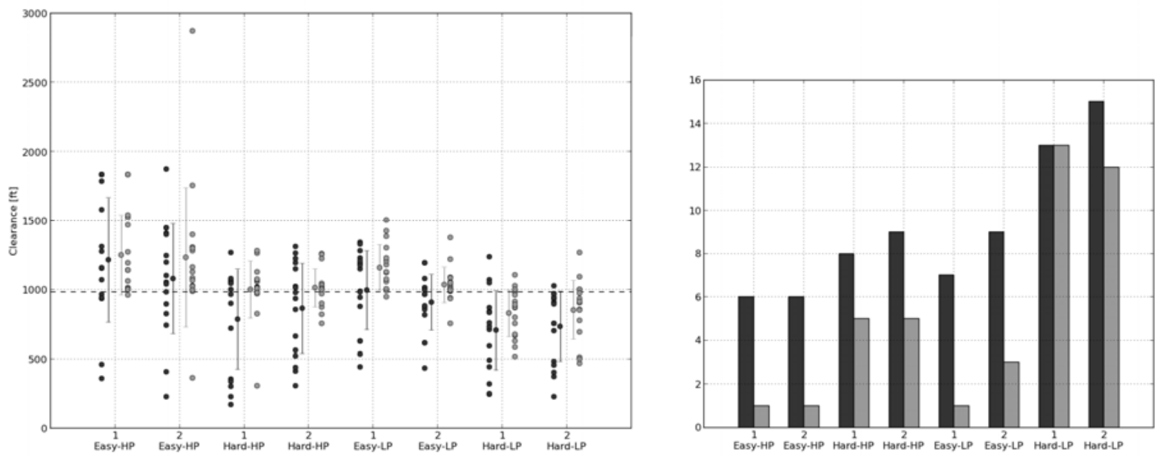


Figure 2: Minimum clearance includeing mean and standard deviation (left) and number of clearance violations (right). Baseline values are black, intentional layer enabled values are gray.

At the end of the experiment each pilot answered four questions about their experience with the intentional constraints. 14 pilots responded positively to the question if the intentional addition makes the terrain avoidance task clearer. The main reasons they give is the increase in situation awareness, better perception of height above the terrain, and a reduced mental load when planning a terrain avoidance maneuver. One pilot stated that the procedural addition makes the task more restrictive. One pilot indicated that the intentional layer was distracting and had not been used.

The question whether the intentional layer changes their strategy, was confirmed by 12 pilots. The majority indicates that the additional information presented enables them to quickly see the lowest regions of the terrain and also provides immediate information about the power required to reach the safe altitude. One pilot also noted that he intentionally put his flight path in the intentional layer because he felt comfortable with a lower clearance and the layer gave him a good indication what pitch angle was required to satisfy this clearance. Of the pilots answering no, two indicated their strategy remained the same but they used the intentional layer as a confirmation of their strategy.

Every pilot, except the one that ignored the intentional constraints, felt that the perceived level of safety increased when the intentional constraints were visible. This mainly happens because the safety margin becomes explicit in the display enabling them to directly assess the risk involved. One pilot noted that adding the procedural constraints *takes the guessing out of flying*. The final question whether the pilots considered the procedural additions useful was answered positively by all but on pilot and is in line with the answers to the previous questions.

Discussion

The main objective of the experiment was to investigate the effect of visualizing intentional constraints in addition to causal constraints helps pilots in making better decisions. Better decisions in the context of this experiment means respecting the minimum safe clearance as much as possible. Analysis of the objective clearance parameter confirm this hypothesis. There is a significant increase in minimum clearance when comparing the intentional addition to the baseline display.

To get more insight in this change in clearance, Figure 2 shows the clearance values of all pilots per condition and the number of pilots that flew below the minimum safe clearance. In the Easy-HP condition all but two pilots fly above the minimum safe clearance when the intentional layer is enabled. In their feedback and through observation it became clear that with sufficient performance and margin pilots will treat the amber band as if it is actual terrain and will avoid it. Only one pilot deliberately ignored the layer and accepted a clearance of less than 500 ft. The second pilot violating the minimum clearance kept his FPV close to the amber layer resulting in a very brief excursion just below minimum clearance.

In the Hard-HP conditions, the same strategy surfaces. Pilots are more inclined to try to meet the minimum clearance with the intentional layer enabled. There are more violations than in the Easy-HP condition, but they are all except one minor violations. By steering into the top of the amber band pilots could make an informed choice about sacrificing a little clearance for a quicker route towards the airport.

In the low performance conditions, the same trends can be observed as in the high performance conditions. The main difficulty in the low performance conditions is that the climb performance significantly decreases during the climb. A number of pilots failed to note that there was not enough margin between the maximum climb performance and the amber region, but even in these cases the amber band shows that they are closer to the top of the minimum clearance and can continue relatively safe.

From the figures it can also be seen that with the baseline display a number of pilots flew with less than 500 ft clearance. For two pilots this was a deliberate choice, the other pilots were mainly unaware of their actual clearance. Not counting these deliberate violations would leave approximately 10 instances where pilots flew below 500 ft. This number is only a third of what Borst, Mulder, and Van Paassen (2010) found in a previous experiment with an EID terrain awareness display. The main reason for this difference is probably the fact that pilots had more freedom to perform an escape maneuver.

Trough the observations, post run feedback of the pilots, and the questionnaire it became clear that the majority of the pilots used the intentional addition to either improve or change their strategy in solving the conflict. The way in which they fit the intentional addition in their strategy can differ but they all indicated that it enhanced their analysis and awareness of the task at hand.

One drawback of the current intentional representation surfaced during the experiment. Once a pilot flies below the minimum clearance, the whole top part of the SVD is filled with the amber color. Once this happens, it is no longer possible to directly perceive the difference between a minor violation close to the top of the intentional layer or a dangerous violation close to the terrain. In the future this could be resolved by using different shades of amber to indicate different clearance levels in the intentional layer.

Conclusion

The main objective of the experiment was to investigate the effect of visualizing intentional constraints in addition to causal constraints helps pilots in making better decisions. Better decisions in the context of this experiment means respecting the minimum safe clearance as much as possible. Analysis of the objective clearance parameter confirm this hypothesis. There is a significant increase in minimum clearance when comparing the intentional addition to the baseline display.

Acknowledgments

The authors would like to thank all pilots participating in the experiment for their time, effort, and professional feedback.

References

Borst, C., Mulder, M., & Paassen, M. M. V. (2010). Design and Simulator Evaluation of an Ecological Synthetic Vision Display. *Journal of guidance, control, and dynamics*, 33(5), 1577-1591.

Hajdukiewicz, J. R., Burns, C. M., Vicente, K.J., & Eggleston, R.G. (1999). Work Domain Analysis for Intentional Systems. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 43, pp. 333-337). DOI: 10.1177/154193129904300343

Rasmussen, J. (1997). Risk Management In A Dynamic Society: A Modeling Problem. *Safety Science*, 27(2), 183-213

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. (1994). *Cognitive Systems Engineering*. Wiley.

Vincente, K. J. (1999). *Cognitive Work Analysis*. Lawrence Erlbaum Associates, Inc.

EVALUATION OF A PERIPHERALLY-LOCATED INSTRUMENT LANDING DISPLAY UNDER DUAL-TASK CONDITIONS

Zachary Spielman, Joseph Vargas, Taleri Hammack, Nathan Bulkley, Roger Lew & Brian P. Dyre
University of Idaho
Moscow, Idaho

Previous research found that a peripherally-located instrument landing system (ILS) embedded in a head-up display (HUD) supported equal or better control of glide-path during simulated approach and landing than the traditional centrally-located MIL-STD ILS. Here, we used a dual-task paradigm to examine whether gains in landing precision with the peripheral ILS are also accompanied by a reduction in mental workload. Participants controlled glide-path during simulated instrument landings while simultaneously performing a secondary task monitoring a head-down engine display for fault states. We varied the type of ILS (peripheral vs. MIL-STD) and assessed mental workload using the NASA-TLX and primary and secondary task performance measures: glide-path errors and engine-fault detection sensitivity, respectively. We found equivalent glide-path errors for the two displays, but the peripheral ILS produced lower subjective estimates of mental workload and significantly less dual-task decrement in engine-monitoring sensitivity, indicating that this display affords effective glide-path control with lower reduced mental demand.

Non-alphanumeric displays incorporating more naturalistic and less symbolic representations of actual flight parameters—known as *virtual displays*—can enhance aircraft pilots' spatial orientation, increase overall performance, and reduce workload (Hettinger, Brickman, Roe, Nelson, & Haas, 1996). Using simulations of approach and landing, Bulkley, Dyre, Lew, & Caufield (2009) showed that an instrument landing system (ILS) using a peripherally-located head-up display (HUD) with virtual symbology composed of moving arrows can provide superior landing precision as compared to the military standard ILS symbology (MIL-STD-1787B) presented in its normal central location within the HUD. Furthermore, Bulkley, Spielman, and Dyre (2011) found that such virtual HUD symbology can be restricted to smaller spatial extents in the far visual periphery to reduce HUD clutter without compromising landing performance. Part of the rational for development of these peripheral displays was that in addition to reducing central visual field load and clutter, the displays should also reduce the mental demands of the ILS. While such reductions in mental workload have been found for peripheral virtual displays for controlling airspeed (Cox, 2000), so far the mental demands of the peripheral ILS HUD have not been examined. Here we compare the mental workload imposed by the peripheral ILS to the traditional MIL-STD ILS using two workload assessment methods: a) subjective estimates measured with the NASA-TLX (Hart & Staveland, 1988) and b) dual-task performance, which required participants to perform a secondary visual engine monitoring task while controlling glide-path under instrument flight conditions (IFC).

Roscoe (1980) estimated that 90% of aircraft control is performed using the central visual field. To accurately read most aviation displays, pilots must directly fixate their gaze upon the display and allocate enough mental resources to interpret the alpha-numeric information presented. This workload is particularly high during approach and landing, and can be exacerbated by pilot fatigue (Hart & Hauser, 1987). A virtual display could present landing information to pilots without taxing overburdened central visual field and attentional resources by taking advantage of automatized orienting and motion coding processes that are particularly robust in peripheral vision. Cox (2000) and Schaudt, Caufield, and Dyre (2002) showed that for simulated flight, a virtual speed error indicator composed of moving fields of arrows projected to the visual periphery provided better flight path and airspeed control than a head-up display (HUD) speed indicator defined by MIL-STD-1787B, while simultaneously lowering subjective workload. These studies clearly demonstrated that the peripheral visual field can process the optical flow created by moving display elements and extract meaningful speed information with less attentional demand and central visual field resources than traditional symbolic displays.

More recently, Bulkley et al. (2009) found that a similar peripherally-located virtual display could be adapted to replace the flight command bars of a traditional ILS display, which are typically presented in the center of the HUD (as defined by MIL-STD 1787B). The ILS command bars instruct pilots to the direction the aircraft needs to be moved to attain proper lateral and vertical alignment with the optimal approach path during landing. The display contains a horizontally-oriented bar that moves up and down to indicate vertical angular errors from the

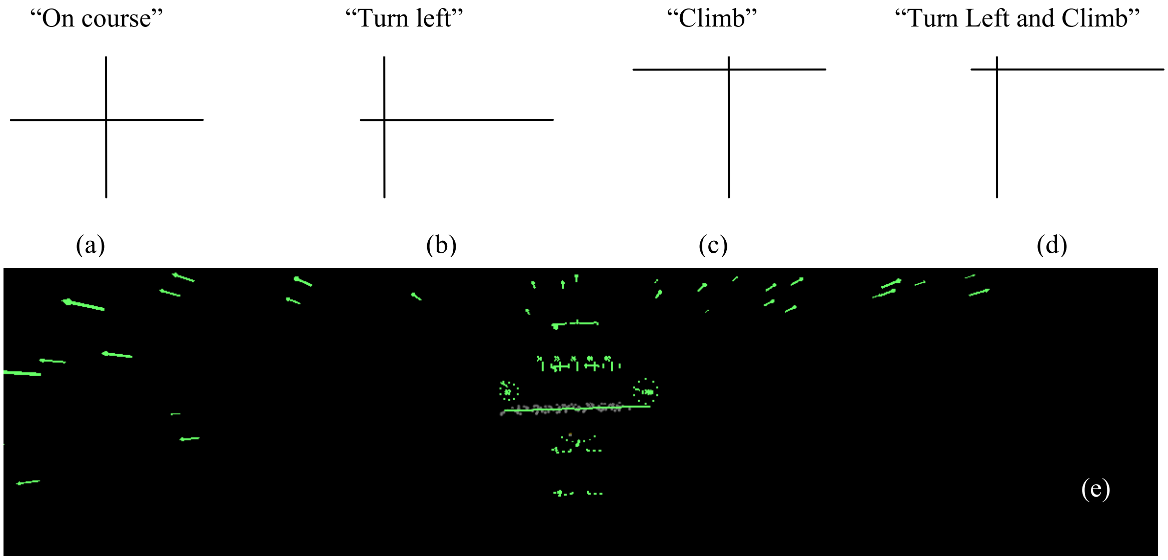


Figure 1. Panels (a)-(d) show the format of MIL-STD 1387B HUD ILS command bars. The four panels show display configurations representing: (a) zero lateral and vertical error, on course, no control correction needed, (b) rightward lateral error, the aircraft needs to turn left to reestablish the optimal approach path, (c) downward vertical error, the aircraft needs to climb to reestablish the optimal approach path, and (d) both downward and rightward errors, the aircraft needs to turn left and climb to reestablish the optimal approach path. Panel (e) shows the peripheral optical flow ILS configuration that would be equivalent to the MIL-STD ILS shown in panel (d), instructing the pilot to "Turn-left and Climb."

ideal glide slope, and a vertically-oriented bar that moves left-to-right to indicate lateral angular errors from the ideal approach path (see Figure 1, panels a-d). When the bars form a symmetrical cross "+" the aircraft is optimally aligned with the ideal approach path. Movements of the bars from the "+" configuration indicate that control inputs are needed to regain the optimal path, and serve as command displays for how the pilot must maneuver the aircraft. In essence, pilots "chase the bars." Bulkey et al. (2009) adapted the peripherally-located virtual HUD format developed for representing airspeed by Cox (2000; see also Schaudt et al., 2002) to an ILS display by simply replacing the centrally-located flight command bars with fields of moving arrows located in the visual periphery (See Figure 1, panel e). In essence, pilots "chase the arrows." The magnitude of lateral and vertical errors is redundantly coded by the speed of the moving arrows and their size, which both increase as errors increase up to a maximum limit. Importantly, size changes occur in discrete steps, creating transient events that can alert a pilot to increasing errors *pre-attentively* (Egeth & Yantis, 1997), even when the displays are presented in the visual periphery. Also, to help reduce display clutter, the arrows disappear completely when errors fall below a minimum threshold.

Here, we are interested in determining whether the peripheral ILS does indeed reduce attentional demand and lower mental workload as compared to the MIL-STD ILS. We report results from one experiment that measured mental demands of the two displays, both objectively, by assessing dual-task performance, as well as subjectively, by administering the NASA-TLX workload questionnaire (Hart & Staveland, 1988). Our primary task required non-pilot participants to fly a simplified simulation of approach and landing at night with poor visibility and significant wind disturbances under instrument flight rules (IFR) using one of the two ILS displays, which we varied within-subjects. We measured glide-path performance—lateral and vertical deviations from glide-path—on the primary task alone and when it was combined with a secondary engine-monitoring task. Performance on the engine monitoring task was assessed as the sensitivity (A') of detection of double engine-faults. Primary and secondary task performance and the unweighted NASA-TLX score or RawTLX (Hart, 2006) served as our measures of mental workload.

Method

Participants

The experiment tested 16 volunteer participants from the undergraduate population of the University of Idaho who received course extra credit as compensation. Participants reported normal or corrected-to-normal acuity and no previous aviation piloting experience. Bulkley et al. (2009) showed that non-pilots using simplified flight controls fly similar flight paths to those flown by pilots for simulated visual approaches under “blackhole” conditions (Gibb, Schvaneveldt, & Gray, 2008), which suggests that pilots and non-pilots are able to perceive their dynamic spatial orientation relative to the runway in a similar manner. Hence, we believe that our sample of non-pilots validly represents the same visual processes that pilots use to control their aircraft during landing.

Apparatus and Stimuli

Visual displays simulating flight were created by a set of five personal computers running ViEWER v2.23 (Dyre, Grimes, & Lew, 2009). One computer served as the simulation host, which coordinated the activity of four graphics channels over a local area network. Three of the graphics channels rendered the main forward view of the environment to three display projectors, which front-projected images with a spatial resolution of 1024 x 768 pixels (H x V) at a refresh rate of 60 Hz onto three large screens arranged as three sides of an octagon with the design viewpoint at the center of the partial-octagon, 1.8 m from the center of each screen. Together, the three main screens subtended 135 x 33.75 degrees of visual angle (H x V).

The fourth graphics channel rendered a head-down display of two pairs of engine status gauges on a 0.17m diagonal LCD screen. The screen was located approximately 1m directly in front of the participant and below the center of three main display screens, and projected a 4 x 3 (H x V) degree image centered at a declination angle of 20 degrees. This position ensured the display was not readable when the eyes fixated the HUD depicted in the forward view, though it remained visible in the periphery. Conversely, when the head-down engine display was fixated, the forward view displayed on the three main screens was visible only in the periphery. The head-down display, shown in Figure 2, consisted of temperature and pressure gauges for two engines. The red status indicators moved throughout the trial according to a unique pseudo-random disturbance determined for each trial. The tic marks labeled “high” and “low” indicated the thresholds at which that engine would be considered to have a critical fault. The combination of movements across the four indicators was programmed to produce ten single faults—only one engine showed a critical fault (or two faults)--and ten double faults—both engines showed at least one critical fault simultaneously. We instructed participants to respond whenever they detected a double fault by pulling the trigger button on the joystick.

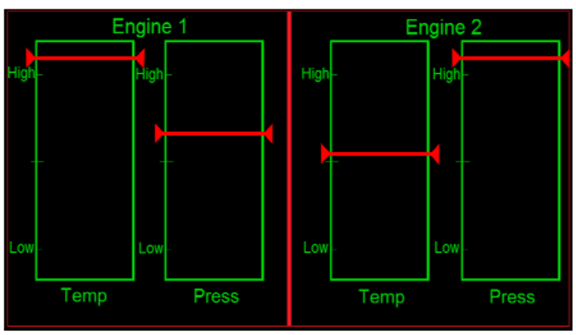


Figure 2. Engine status indicators displayed on the head-down monitor used for the secondary engine fault detection task. These indicators are consistent with a “double-fault” requiring

The simulated environment displayed on the three main screens consisted of a large island extending 288,000 m in width (x) and 72,000 m in depth (z). A 30m wide runway started at the center of this island [x, y, z] = [0, 0, 0] and extended 950 m in depth [0, 0, -950]. Green, white, and red lights were spaced every 50m along the sides of the runway in a standard runway lighting configuration. The runway was texture-mapped with a pavement texture and the surrounding ground surface was texture-mapped with a high spatial frequency seamless texture resembling dirt and grass. At the far end of the runway the ground began to slant upward at 5 degrees for 18,000m before leveling off at a maximum elevation of 1600 m. The virtual environment simulated ambient lighting equivalent to a foggy, overcast, moonless night such that the unlit ground became barely visible at an altitude of 160 m and the runway marker lights became visible at a distance of approximately 2000m..

The simulated view from the vehicle was open, no windscreen frame or other artifacts were used to create a simulated cockpit. The vehicle started 9889m short of the runway at an altitude of 720m, aligned laterally with the center of the runway and moving forward on a level trajectory at 45 ms^{-1}. The target glide-slope angle of 7.3 degrees was defined via two invisible waypoints. The first was located at the vehicle starting point [0, 720, 9889], and the second was centered on the runway, 2m from its near end [0, 0, -2].

Simulated lateral and vertical wind disturbances were produced by translating the vehicle along the lateral (x) and vertical (y) axes independently. Each disturbance was defined by a sum-of-sines with five prime frequencies. Amplitudes were chosen such that the maximum acceleration did not exceed 1 G (9.8 ms^{-2}). The disturbance frequencies for the x- and y-axes were 0.055, 0.085, 0.145, 0.185, 0.215, and 0.035, 0.065, 0.115, 0.155, 0.205 Hz, respectively; amplitudes were 10.47, 2.34, 1.37, 1.08, 0.93 and 25.85, 7.49, 2.39, 1.32, 0.75 m, respectively.

Since participants were non-pilots, simplified flight controls were used to control the lateral and vertical positions of the simulated aircraft. Left-right movements of a CF-F-16 joystick with the right hand controlled lateral velocity with a first-order transfer function with gain = 25 ms^{-1} at maximum stick deflection and exponential lag constant = 1.0 s^{-1}. For simulation of banking while moving laterally, left-right stick movements also produced a zero-order roll with a gain = 15° at maximum deflection and exponential lag constant = 0.5 s^{-1}. Movements of a CF Pro Throttle configured as a first-order controller with gain = 25 ms^{-1} and exponential lag constant = 1.0 s^{-1} with the left hand controlled vertical velocity. Forward movement of the throttle caused the vehicle to go up (simulating more thrust); backward movement caused the vehicle to go down (less thrust).

Similar to Bulkley et al (2009), monochromatic green HUDs were superimposed over the terrain and environmental objects on the display and included the following indicators from MIL-STD-1787B: aircraft pitch reference symbol, climb/dive marker, climb/dive ladder, airspeed indicator, target airspeed indicator, altitude indicator, heading indicator, and bank indicator. Two ILS formats were implemented. One ILS format was the MIL-STD 1787B represented as vertical and horizontle bars refferred to as course deviation and vertical deviation indicator bars respectively (hereafter referred to as *MIL-STD*). The other ILS was implemented within the HUD as a peripherally-located virtual display of fields of moving arrows (hereafter referred to as *Peripheral OF*). Fields of moving arrows randomly arranged within a volume of space appeared in the upper, lower, left or right peripheral areas of the HUD to provide control commands to overcome lateral and vertical deviations from the optimal flight path (see Figure 1). The visual fields subtended by the arrows were equivalent to the far periphery condition of Bulkley et al (2011), since they found no decrement in landing precision when the field of arrows was constrained to the far periphery. Similar to the MIL-STD ILS, the Peripheral OF ILS used a command format that informed pilots which direction to move to correct their course—in effect participants needed to "chase the arrows" to maneuver toward the optimal glide path. The arrows coded lateral and vertical deviation error magnitudes redundantly using both size and speed of movement. Zero course error resulted in zero speed and size--nothing was displayed. Small course errors resulted in small arrows moving slowly. As course errors increased, the speed of the arrows increased proportionately, and the size of the arrows increased in step-wise increments to produce sudden size-changes that naturally captured attention.

Experimental Design and Procedure

A 2 x 2 x 3 within subject factorial design tested the effects of ILS format (MIL-STD, Peripheral OF), block (1 or 2), and task (Single-Task Engine Monitoring, Single-Task Landing, Dual-Task) on glide-path errors (vertical and lateral), engine-monitoring performance, and subjective workload. Testing occurred over two 90-minute sessions, with the ILS format blocked across session. We tested half of the participants first with the MIL-STD format then the Peripheral OF; the other half of participants received the reverse order. Each testing session started with verbal instructions on the physics of flight, relation of control movements to aircraft movements, a review of all display elements, and a description of the engine monitoring task. Following this training, half the participants were first tested on the engine monitoring single task, then the landing single task, while the other half received the reverse order. We tested the engine monitoring single-task across three trials, and the first trial was considered training and not included in the data analysis. To become acquainted with the flight controls for the landing task, participants performed a day-time landing with no ILS using visual flight rules (VFR). Following this, the experimenter demonstrated an IFR landing using the ILS assigned to the participant for that session to familiarize the participant with the ILS format and night-time conditions. We then tested the landing single-task across six trials, and the first four trials were considered practice and not included in the data analysis. At the end of each session, all participants performed three trials of the dual-task condition, of which the first trial was considered practice and not included in the data analysis. Each trial lasted approximately 3 minutes and 40 seconds and participants were able to complete all 14 trials, plus instructions and debriefing within two sessions scheduled on successive days. At no time were the participants instructed to give priority to either of the tasks. After each block of trials for a task, participants completed the NASA TLX.

Results

Glide-path (Primary) Task Performance

Altitude errors were defined as the difference between actual altitude and the target altitude defined by the linear glide slope at a particular point along the approach. Lateral errors were defined as the difference in position of the simulated aircraft relative to the runway centerline. We computed the root-mean-squared (RMS) of these errors for each trial to represent overall error glide-path errors.

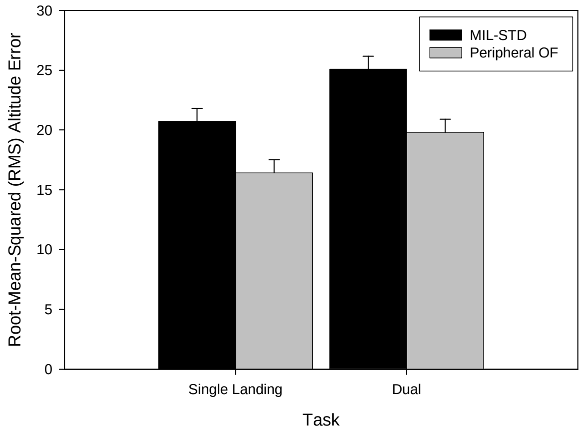


Figure 3. Mean RMS calculated for all participant vertical deviation error. Error bars report average standard error across participants for each category. Error bars represent within-subjects standard errors of the mean calculated from the MSE term of the task x display interaction.

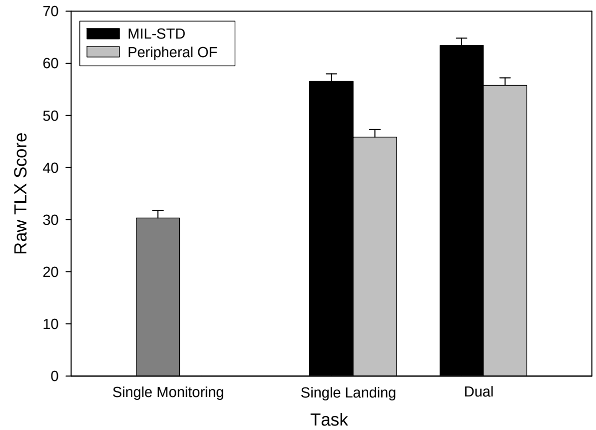


Figure 4. Raw TLX subjective measure workload. Error bars report average standard error across participants for each category. Error bars represent within-subjects standard errors of the mean calculated from the MSE term of the task x display interaction.

RMS lateral and altitude errors were each analyzed using a 2 x 2 x 2 x 2 factorial mixed analysis of variance (ANOVA) with ILS configuration (Peripheral OF, MIL-STD), task (single, dual), and block (1-2) as within-subjects factors and order of ILS configuration as a between-subjects factor. Though the pattern of RMS Altitude error in Figure 3 suggests main effects for ILS and Task, these trends were non-significant, $F(1, 14) = 3.733$, $p = .074$, $MSE = 208.14$, and $F(1, 14) = 3.613$, $p = .078$, $MSE = 50.77$, respectively. No reliable effects were found for lateral error.

Engine Monitoring (Secondary) Task Performance

We computed sensitivity in dual engine fault detection as A' and analyzed these values using 3 x 2 x 2 mixed ANOVA with Task-Display Condition (Single Task Engine Monitoring, Dual-Task Peripheral OF, Dual-Task MIL-STD) and block as the within-subjects factors and order as a between-subjects factor. We found a significant main effect of task-display condition, $F\ (2, 28) = 24.29$ $p = 0.001$, $MSE = .002$. According to this effect, detection performance was best in the Single-Task condition, followed by the Dual-Task Peripheral OF condition, with worst performance in the Dual-Task MIL-STD condition (see Table 1).

Subjective Workload

Data for two participants was lost due to technical glitches. We computed Raw TLX scores for the resulting 14 participants by taking the average score across the six subscales and analyzed these scores using a 2 x 2 x 2 x 2 factorial mixed analysis of variance (ANOVA) with ILS configuration (Peripheral OF, MIL-STD), task (single, dual), and block (1-2) as within-subjects factors and order of ILS configuration as a between-subjects factor. We found significant main effects of Display and Task condition, $F\ (1, 12) = 6.13$ $p = 0.033$, $MSE = 164.35$ and $F\ (1, 12) = 11.88$ $p = 0.006$, $MSE = 71.31$ respectively. The pattern of means, listed in Table 1 and shown in Figure 4, demonstrate that the dual-task condition was significantly more demanding than the single task conditions and that higher demand occurred with the MIL-STD versus the Peripheral OF. We also found a significant ILS-Order interaction, $F\ (1, 12) = 16.73$ $p = 0.002$, $MSE = 164.35$, which showed overall lower workload ratings for testing on Day 2 as compared to Day 1.

Discussion

Our peripheral optical flow ILS clearly out-performed the traditional MIL-STD ILS format in providing precise glide-path control with lower cognitive demand. The peripheral ILS provided at least equivalent performance to the MIL-STD ILS while significantly improving performance on a secondary visual monitoring task, indicating that participants experienced less resource demand while landing with the peripheral ILS. Significantly lower subjective assessments of mental workload for the peripheral ILS confirmed that participants were even aware they experienced lower workload with the peripheral ILS.

Table 1

Glide-path Root-mean-squared (RMS) Errors, Detection Sensitivity (A'), and RAW-TLX Scores by Task/Display Condition

Task/Display	RMS Lateral Error (m)	RMS Altitude Error (m)	Detection A'	RAW-TLX Composite
Single-Task Engine Monitoring	-	-	.964 (.009)	35.01
Single-Task Landing				
MIL-STD ILS	9.60 (0.52)	20.42 (1.09)	-	56.53 (1.44)
Periperhal OF ILS	9.22 (0.52)	16.18 (1.09)	-	45.83(1.44)
Dual –Task Landing				
MIL-STD ILS	9.54 (0.52)	23.5 (1.09)	.887 (.009)	63.40 (1.44)
Periperhal OF ILS	9.03 (0.52)	17.88 (1.09)	.920 (.009)	55.76(1.44)

Note. Numbers in parentheses represent within-subject standard errors (SEs) of the mean. For RMSE these SEs are based on the Display x Task x Subject mean square error (MSE). For detection A' these SEs are based on the Task for the comparison of display configurations.

These results have important implications for the design of HUDs. Clearly, the peripheral visual field is a potentially important visual resource that is underused with current HUDs and may be a particularly valuable resource for processing flight parameters related to spatial orientation, such as ILS control of landing approaches. Peripheral virtual displays of ILS commands appear to provide our non-pilot participants with information in a more natural, pre-attentive manner that lessens resource demand, central visual field load, and display clutter, while still affording performance that is equal to or better than traditional ILS displays. Further research is needed to determine whether these benefits generalize to actual pilots.

References

Ashford, R. (1998). *Approach and Landing Fatal Accidents 1980–1996*, Safety Analysis Report No. SA-98-002, Civil Aviation Authority, London.

Boeing Commercial Airplane Group. (2002). Statistical summary of commercial jet airplane accidents, worldwide operations, 1959-2002. Retrieved December 8,2002, from http://www.boeing.com/news/techissues/pdf/statsum.pdf

Bulkley, N., Dyre, B. P., Lew, R., & Caufield, K. (2009). A Peripherally-located Virtual Instrument Landing display affords more precise control of approach path during simulated landings than traditional instrument landing displays. *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society,* 31-35. San Antonio, TX: HFES

Bulkley, N. K., Spielman, Z., & Dyre, B. P. (2011) Peripherally-located Virtual Instrument Landing Displays. In *Proceedings of the 16th International Symposium on Aviation Psychology (ISAP)*, Dayton.

Cox, E. M. (2000). The effect of a virtual indicator on speed control. (Unpublished Master's Thesis, University of Idaho).

Dyre, B. P., Grimes, J. G., & Lew, R. (2009). ViEWER 2.23: A Virtual Environment Workbench for Education and Research. *http://www.webpages.uidaho.edu/~bdyre/viewer.htm.*

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48,* 269-297.

Gibb, R., Schvaneveldt, R., & Gray, R. (2008). Visual misperception in aviation: glide path performance in a black hole environment. *Human Factors, 50*(4), 699-711.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 904-908). Santa Monica, CA: Human Factors & Ergonomics Society

Hart, S. G., & Hauser, J. R. (1987). Inflight application of three pilot workload measurement techniques. *Aviation, Space, & Environmental Medicine, 58*(5),402-410.

Hart, S., G. & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North Holland.

Hettinger, L. J., Brickman, B. J., Roe, M. M., Nelson, W. T., & Haas, M. W. (1996). Effects of virtually-augmented fighter cockpit displays on pilot performance, workload, and situation awareness. *Proceedings of the Human Factors Society 40th Meeting*, 30-33. Philadelphia, PA: HFES.

Kaber, D. B., Alexander, A. L., Stelzer, E. M., Kim, S., Kaufmann, K., & Hsiang, S. (2008). Defining and psychological modeling of perceived clutter in advanced cockpit displays. *Aviation Space and Environmental Medicine, 79,* 1007-1018.

Khatwa R., Collins R., & Helmreich, R.L. (1999). Data acquisition and analysis working group final report. *Flight Safety Digest,*1-77.

Roscoe, S. (1980). *Aviation Psychology.* Ames, IA: Iowa St. University Press.

Schaudt, W.A., Caufield, K.J., and Dyre, B.P. (2002). Effects of a virtual speed indicator on guidance accuracy and eye movement control during simulated flight. *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society,* 318-322. Baltimore, MD: HFES

Wickens, C. D. (1991). Processing resources and attention. In D. Damos (Ed.), *Multiple task performance.* London: Taylor & Francis.

EXPERIMENTAL EFFECTIVE INTENSITY OF STEADY AND FLASHING LIGHT EMITTING DIODES FOR AIRCRAFT ANTI-COLLISION LIGHTING

Chris Yakopcic, John Puttmann, Benjamin R. Kunz, Ph.D., Clara Ang, Ashley McPherson, David Santez, Matt Donovan, John Skarzynski, and Joshua Trick
University of Dayton
Dayton, Ohio

Andrew Mead, Ph.D. and Nelda Milburn, Ph.D.
Federal Aviation Administration
Civil Aerospace Medical Institute
Oklahoma City, Oklahoma

Nazih E. Khaouly
Federal Aviation Administration
Transport Airplane Directorate
Airplane and Flight Crew Interface Branch
Renton, Washington

Research was conducted to determine the effective intensity of flashing lights that incorporate light-emitting diodes (LEDs). LEDs require less power and have the ability to flash without the addition of moving parts. Compared with incandescent bulbs, however, LEDs yield a different spectral output and a different intensity profile when flashing. To determine the effect of these differences on a viewer's ability to detect the light, we examined LEDs to determine if they can successfully replace legacy technologies/assemblies on aircraft. The LED was displayed to naïve subjects to establish visibility thresholds using an automated system to drive the LED with variable intensity and duration. Experimental data were examined to determine which model for effective intensity (Allard, Modified Allard, or Blondel-Rey) is most appropriate for LEDs. Each of the methods was found to be applicable dependent upon the system being considered. Use of the Blondel-Rey method produced acceptable but conservative results.

In recent years, designs for aircraft anti-collision lights have incorporated LEDs because of their low power consumption, longer operating life, and the ability to flash without moving parts. However, the spectral output of an LED is significantly different than that of existing technologies. A white incandescent bulb produces a yellowish output and most white LEDs produce a white that has a much stronger blue component. Also, the intensity profile, or pulse shape of a flashing LED is rectangular and much longer compared with a xenon flash tube having very short pulses, and different from a rotating incandescent bulb producing some variation of a rounded pulse shape.

The research presented in this paper evaluated several formulae (the Blondel-Rey, the Allard, and the Modified Allard) to find an acceptable and practical measurement technique for describing the effective intensity of flashing LED lights, such as aircraft anti-collision light system, that incorporate LED technologies. The method incorporated in Title 14 of the Code of Federal Regulations (CFR) Part 25, §25.1401(e) has been adequate for use with xenon flash tube technology, but may be inappropriate for anti-collision lights using LED technology. According to existing standards, the effective intensity for an anti-collision light must be determined by the Blondel-Rey formula, specified in CFR §23.1401, 25.1401, 27.1401 and 29.1401. These new LED anti-collision lights use various pulse durations, pulse shapes, and groupings of pulses to generate the intensity and flashing characteristics required in §25.1401. Because data have indicated that the Blondel-Rey metric may underestimate the effective intensity of these flashing lights with complex pulse patterns, the Blondel-Rey equation may be inappropriate for determining the effective intensity of anti-collision lights using LEDs (Federal Aviation Administration, 2009).

Effective intensity is defined by International Commission on Illumination (CIE) as "luminous intensity of a fixed (steady) light, of the same relative spectral distribution as the flashing light, which would have the same luminous range (or visual range in aviation terminology) as the flashing light under identical conditions of observation." The currently-used FAA standard for qualifying the effective intensity of flashing lights was first proposed by Blondel and Rey in 1911. That experiment involved subjects viewing a lamp housed in a contraption with a rotating disc that created a flashing light effect. Since then, the technology used to develop aircraft anti-collision lights has changed considerably, yet the equation proposed by Blondel and Rey is still used as the standard.

The Blondel-Rey equation has been evaluated both experimentally (1911) and analytically (Ohno & Couzin, 2002) for use with LEDs. In addition, other models have been proposed as alternatives for measuring the effective intensity of a pulsed light source such as the Allard method (Allard, 1876) and the form-factor method (Schmidt-Clausen, 1968). Ohno and Couzin (2002) conducted a theoretical study of these models and proposed the

Modified Allard method as a more accurate alternative for multi-pulse flashing lights. Multi-flash signals are advantageous, as they help the observer to better judge the distance and location of the light signal (Mandeler & Thacker, 1986). Experimental effective intensity of multi-flash signals is studied (Mandeler & Thacker, 1986) using very short xenon strobe pulses. Our previous work in this area used the same apparatus described below to determine experimental effective intensity of a single-pulsed LED (Yakopcic et al., 2012). Those results showed that using a randomized distribution of LED signals resulted in a stronger correlation to the Blondel-Rey equation than when presenting the signals using a method of limits approach.

Methods for Modeling Effective Intensity

The three models for effective intensity examined in this paper include the Blondel-Rey equation, the Allard Method, and the Modified Allard method. The Blondel-Rey equation is shown in Eq. (3), where *I(t)* is the intensity profile of the pulse output from the LED, and *a* is the visual time constant that was experimentally determined to be 0.2s by Blondel and Rey. For a square pulse, t_1 and t_2 are the rising and falling edge of the pulse respectively. The value I_{eff} is the effective intensity of the light pulse in question.

$$I_{\text{eff}} = \frac{\int_{t_1}^{t_2} I(t)dt}{a+(t_2-t_1)} \quad (1)$$

The Allard method is based on a convolution of the intensity profile of the pulsed LED with the visual impulse response function as defined by Allard. The equations for the Allard method are shown in Eqs. (2) and (3), where *I(t)* again represents the intensity profile of the light pulse. The function *i(t)* is the convolution of *I(t)* and the visual impulse response *q(t)* as defined in Eq. (3). Convolution refers to the mathematical operation denoted by the ⊗ symbol in Eqs. (2) and (4). The visual time constant *a* is also set to 0.2s in this equation (Ohno & Couzin, 2002), and the effective intensity is defined as the maximum of *i(t)*.

$$i(t) = I(t) \otimes q(t) \quad (2)$$

$$q(t) = \frac{1}{a} e^{-\frac{t}{a}} \quad (3)$$

The Modified Allard method (Ohno & Couzin, 2002) was first developed by optimizing the constants in the visual impulse response function given in Eq. (5) so that the result of this method closely matched the Blondel-Rey equation for single rectangular pulses. Further theoretical analysis discovered the *q*(*t*) function that perfectly matches the results of Blondel-Rey for rectangular pulses, shown in Eq. (6).

$$i(t) = I(t) \otimes q(t) \quad (4)$$

$$q(t) = \frac{w_1}{a_1} e^{-\frac{t}{a_1}} + \frac{w_2}{a_2} e^{-\frac{t}{a_2}} \quad (5)$$

$$q(t) = \frac{a}{(a+t)^2}; \, t \geq 0 \cdot \quad (6)$$

Experimental Method

Subjects

Participants were recruited through email lists of enrolled students and with posters hung in campus buildings. Prior approval for all procedures and use of human subjects was obtained from the University of Dayton (UD) Institutional Review Board. Informed consent was obtained prior to participation and subjects were free to withdraw from the project without consequence at any time. Three different experiments were conducted that differed in the pulse type displayed. These included (1) a single rectangular pulse, n=36; (2) a 2 pulse multi-flash signal, n=21; or (3) a 4-pulse multi-flash signal, n=15. In a one-hour testing session, only one of these three signal options was utilized. Each test subject completed a demographics form for general data regarding eyesight. A visual acuity test was administered and only subjects with 20/30 vision or better were utilized for the experiments. The subjects were mainly college students, and the subject age was in an 18-25 range with about 50% males and 50% females. Subjects were rewarded with a $20 voucher for the UD Bookstore upon completing the experiment.

Apparatus and Procedure

A MATLAB (Mathworks) program was developed that automatically controlled the LED intensity and pulse width that was observed by the test subjects. The MATLAB script was capable of controlling the pulse width to a resolution of 2 milliseconds. The MATLAB script fully automated the testing procedure for displaying the LED and collecting the user responses using input obtained from a “yes” button and a “no” button placed below the subject’s right and left hand respectively that corresponded to whether or not a light was seen. A single trial consisted of a 5-second presentation and answer interval and a 2-second between-trial interval. A total of 539 trials were presented at random. These trials consisted of 7 different pulse widths (see Figure 1) at 11 different intensities (see Table 1), each shown 7 times. Preceding the trials, a dark adaption and pretest was administered. This consisted of a 7-minute dark adaption period, followed by a 16-trial pre-test, and then another 8 minutes of dark adaption.

Subjects were seated 50 feet directly in front of the LED apparatus. The LED was white with chromaticity values of (x=0.301, y=0.293) on the 1931 CIE chromaticity diagram. Black Curtains were hung to remove the possibility of reflection from the walls. The LED was housed in a wooden, matte black box with circular baffles to reduce the scattering of the observed light. A headrest was used to ensure the subjects were looking in the direction of the LED. The LEDs required about 3mA to display a stable signal, which provided too much intensity for these experiments. To obtain appropriate intensity values considered to encompass each test subject’s threshold based on a priori testing, neutral density filters were used to reduce the light output from the LED to the order of micro-candelas (µcd). Our system was capable of driving the LEDs with a maximum current of about 30mA. Therefore, to maximize the range of intensities that could be presented, two identical LEDs were used side by side. Only one LED was presented within a single trial, although different neutral density filters were applied to each LED. For the single pulse test, the neutral density filters differed by a factor of 10. Given that each LED also has an output range of about a factor of 10, this allowed for a set of intensities where the strongest intensity was about 100 times that of the weakest (see Table 1, Single Pulse Test), hence permitting a wide range of brightness for a more accurate determination of each subject’s threshold.

Table 1.
The values in the table show the peak luminous intensity of each rectangular pulse after considering neutral density filters for each experiment. Reduction in transmission through filters is linear for all visible wavelengths.

Experiment 1 n=36 Single Pulse Test LED Intensities (µcd)		**Experiment 2 n=21** 2-Pulse Multi-Flash Test LED Intensities (µcd)		**Experiment 3 n=15** 4-Pulse Multi-Flash Test LED Intensities (µcd)	
LED 1	LED 2	LED 1 (Steady)	LED 2 (2 Flash)	LED 1 (Steady)	LED 2 (4 Flash)
0.00	0.00	0.00	0.00	0.00	0.00
0.02	LED Not Used	0.06	0.39	0.06	0.20
0.34		0.08	0.51	0.08	0.26
LED Not Used	0.38	0.10	0.60	0.10	0.30
	0.81	0.20	1.28	0.20	0.64
	1.23	0.31	1.94	0.31	0.97
	1.62	0.41	2.57	0.41	1.29
	2.00	0.50	3.17	0.50	1.59
	2.36	0.59	3.74	0.59	1.88
	2.72	0.68	4.32	0.68	2.16
	3.06	0.77	4.85	0.77	2.43

In addition to the variable intensities, 7 different pulse widths were used; 5s, 1s, 500ms, 250ms, 100ms, 50ms, and 20ms. The 5-second light represented a steady state signal because all methods for calculating effective intensity showed that at a pulse width of 5 seconds resulted in an effective intensity within 5% of the intensity of a true steady state light. The steady state (5-second) signal is required as the reference to accurately determine the effective intensity. Figure 1 shows the pulse width of the normalized LED signal and response interval for experiments 1 and 2. Experiment 3 has a similar pattern to experiment 2, although each trial width contains 4 pulses. The 5-second, steady state signal directly overlaps the response interval in each case. Each of the multi-flash tests (experiments 2 and 3) required a 5 second steady state signal for comparison, but the small multi-flash signals were significantly more difficult to detect when compared to the steady state signal. As a solution, the steady state and the multi-flash signal were each presented from a different LED. Each LED was identical but different neutral density filters were applied to each. The intensities for the 2-flash and 4-flash multi-pulse experiments (see Table 1).

Results

For each test subject, a visibility threshold was calculated for each of the 7 pulse widths. This was done by applying a logistic regression to collected data and determining the point where the probability of detection was at 50%. The plots in Figure 2 display a data set from a single subject. The visibility thresholds obtained from all test subjects were used to determine a mean visibility threshold for each pulse width in each of the three experiments. To determine the experimental effective intensity using these data, the mean steady state threshold was divided by the mean threshold for each of other six pulse widths. The results for the single pulse experiment were based on 36 subject datasets, the results for the 2 pulse experiment were based on 21 subject datasets, and the 4 pulse experiment was based on 15 subject datasets.

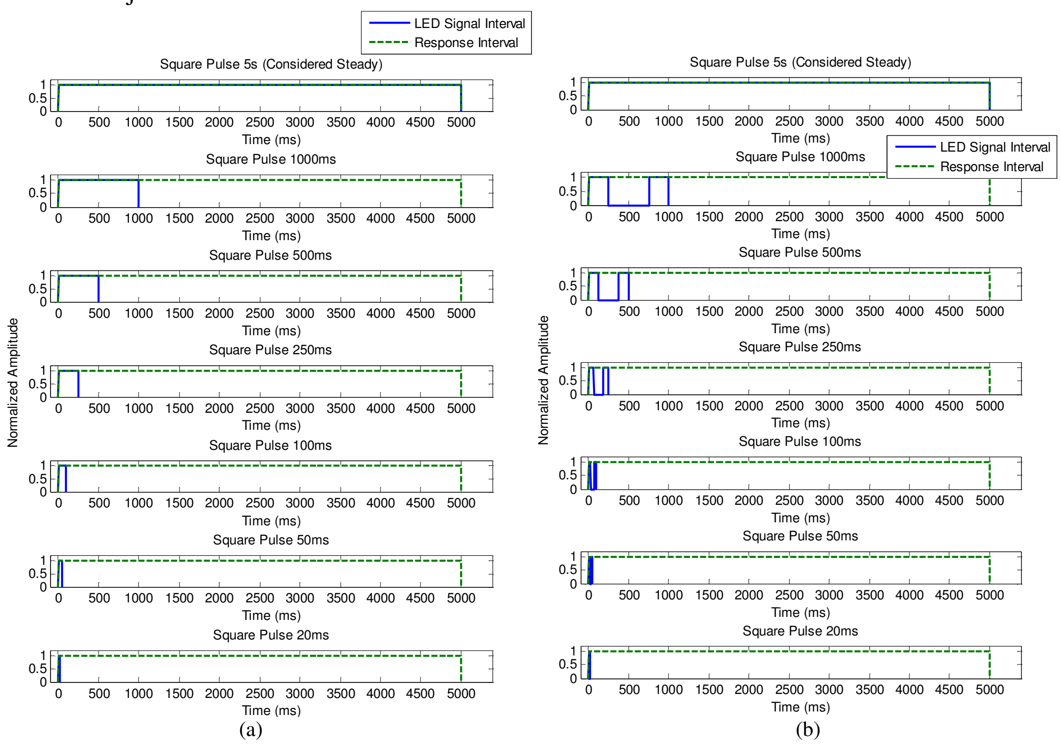


Figure 1. Normalized LED signals and response intervals for (a) the single pulse experiment and (b) the 2 pulse multi-flash experiment.

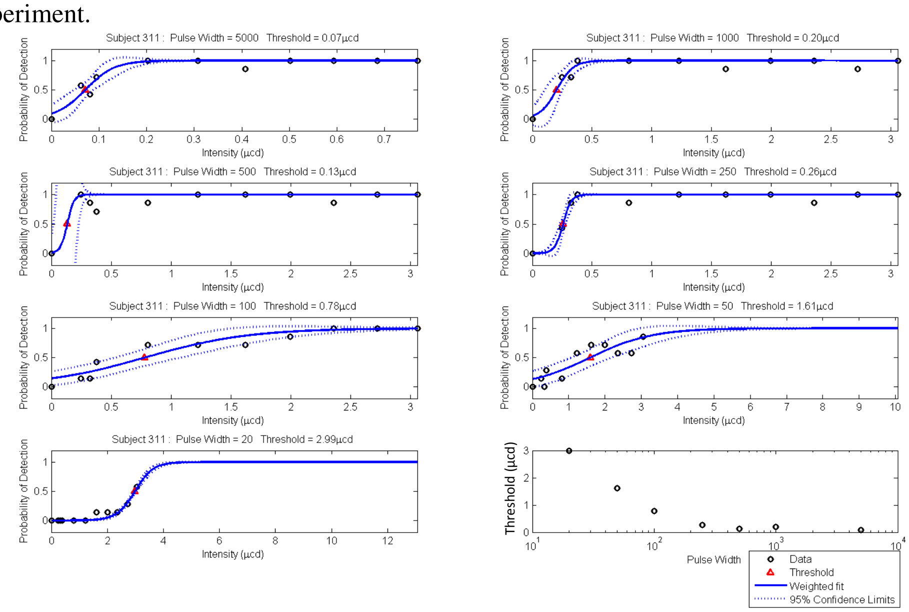


Figure 2. Sample of individual subject data.

Figures 3, 4, and 5 show how the experimentally determined effective intensities relate to the theoretical effective intensity equations including the Blondel-Rey, Allard, and Modified Allard methods. It should be noted that the multi-flash experiments were compared to the Blondel-Rey-Douglas equation (Douglas, 1957) as opposed to Blondel-Rey equation because the Blondel-Rey-Douglas equation (described by Ohno and Couzin (2002)) provides a slight modification to the Blondel-Rey equation that allows for more accurate modeling of multi-pulse signals. Figures 3, 4, and 5 correspond to the single pulse, 2-pulse, and 4-pulse experiments, respectively. Figure 3 shows that the best representation of the data collected would appear to be the Allard method. Although, Figures 4 and 5 both show that the multi-flash experimental data are more closely related to the Modified Allard method.

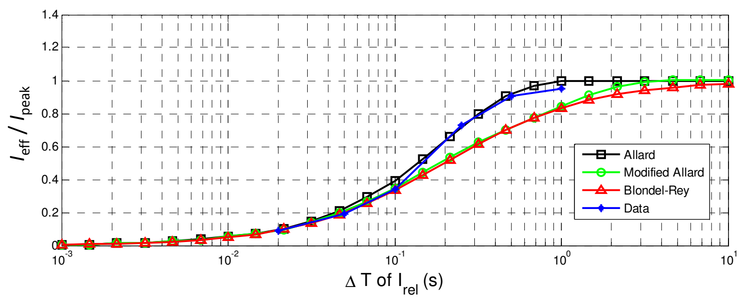


Figure 3. Experiment 1: Comparison of data from the single pulse experiment and the models for determining effective intensity (I_{peak}=Peak signal intensity).

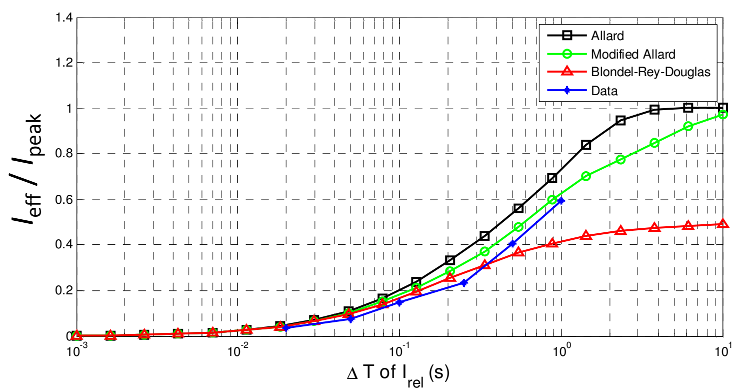


Figure 4. Experiment 2: Comparison of the 2-pulse experimental data and predicted results.

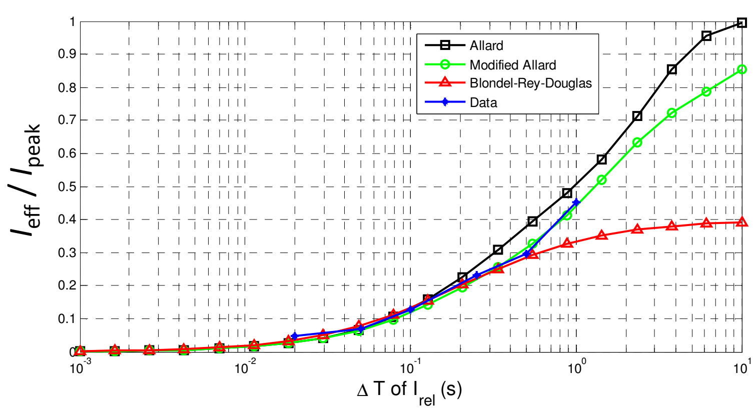


Figure 5. Experiment 3: Comparison of the 4-pulse experimental data and predicted results.

Conclusion

These results suggest that different models for effective intensity are required when using different LED signal patterns. When a single pulse flash is being used, it would appear that the Allard method is most appropriate to model effective intensity, although the data collected from the multi-flash experiments seems to suggest that the Modified Allard method is the best choice to match experimental results. In all cases, use of the Blondel-Rey method requires a higher actual intensity to produce a calculated effective intensity to be perceived as equivalent to the experimental results. Therefore, continued use of the Blondel-Rey method can be considered a conservative approach, but use of Modified Allard and Allard methods, as appropriate, could result in reductions in lighting component mass and/or energy while ensuring performance equivalently. Changes to the visual time constant of 0.2s could possibly make the Blondel-Rey method more appropriate for multiple flash cases when using LEDs, and this could be evaluated in future research.

Acknowledgments

Research reported in this paper was conducted under the Flight Deck Program Directive / Level of Effort Agreement between the Federal Aviation Administration Headquarters, and the Aerospace Human Factors Division (AAM-500) of the Civil Aerospace Medical Institute and is sponsored by FAA ANM-111 and supported through the FAA NextGen Human Factors Division, (ANG-C1). The authors of this paper acknowledge the support of Dr. Yoshi Ohno of the National Institute of Standards and Technology and Loran Haworth of the Federal Aviation Administration. Also the authors thank Professors John Hageman and Dr. Phil Doepker for their insight and academic support. The authors also thank a number of students that dedicated time to this project, including Adelheid Heckel, Philip Kloos, Joseph Martin, and Fangzhou Sun.

References

Allard E., (1876). Mémoire sur l'intensité et la portée des phares", 62-73, Imprimerie Nationale, Paris.

Blondel, A., & Rey, J. (1911). Sur la perception des lumieres breves a la limite de leur portee. Journal de physique thtorique et appliqute., 1, 530-550.

CIE Commission Internationale de I'Eclairage S 017/E:2011, ILV: International lighting vocabulary.

Code of Federal Regulations, Title 14, Parts 23, 25, 27, and 29. Washington, DC: U.S. Government Printing Office, 2011.

Douglas, C.A., (1957). Computation of the effective intensity of flashing lights, Illuminating Engineering, N.Y. Vol. LII.No 12, pp.641-646, 1957.

Federal Aviation Administration, (2009) Internal Document, ANM-111, Safety TCRG R&D Requirement, Effective intensity of flashing lights.

Mandeler, M. B., Thacker, J. R., (1986). A method of calculating the effective intensity of multiple-flick flashtube signals, U.S. Coast Guard Publication CG-D-13-86.

Ohno, Y., & Couzin, D. (2002). Modified Allard method for effective intensity of flashing lights. In Temporal and spatial aspects of light and olour perception and measurement, Expert Symposium. Veszprem, Hungary: Commission Internationale de L'Eclairage.

Schmidt-Clausen, H. J., (1968). Über das Wahrnehmen verschiedenartiger Lichtimpulse bei veränderlichen Umfeldleuchtdichten (Concerning the perception of various light flashes with varying surrounding luminances), Darmstadt Dissertation D17, Darmstadt University of Technology, 1968.

Yakopcic, C., Puttmann, J., Kunz, B. R., Holleran, M., Wingeier, B., Hashemi, A., Stapp, K., (2012). Human perception of light-emitting diodes for aircraft anti-collision lighting, International Symposium for Aviation Psychology.

AIR TRANSPORT INCIDENT AND ACCIDENTS CAUSED BY CREW SITUATION AWARENESS ERRORS

Robert L. Arnold
Denver, Colorado

This paper examines the various aircrew situation evaluation errors that resulted in one near fatal commercial airline incident and two fatal commercial airline accidents. The incident and accidents involved events that distracted the aircrews from a timely and accurate evaluation of the emergency situations that confronted them. Summary narratives of the incident and each of accidents are provided. Each emergency situation is examined to identify the distractions that caused the crews to misinterpret the nature of the emergencies they faced. Situation awareness as a construct is briefly reviewed. Recommendations for situation awareness training for aircrews are made. Recommendations are also made for improvements in aircraft crew alerting systems and operating procedures.

Situation awareness (SA) as a scientific construct has emerged in human factors research. Durso and Sethumadhavan (2008) provide an overview of SA research and a discussion (p. 444) of the SA component (e.g., *situation assessment*) that I address in this paper. The near fatal incident and two fatal accidents examined in this paper all resulted from inaccurate situation assessments. These inaccurate situation assessments were caused by distracting events that were not immediately relevant to the actions the aircrews had to take to resolve the critical situations they faced. The near fatal incident involved United Airlines Flight 863, from San Francisco, California to Sydney, Australia on June 28, 1998. The fatal accidents include Helios Airways flight HCY522, from Larnaca, Cyprus to Athens, Greece on August 14, 2005 and Air France flight AF 447, from Rio de Janeiro, Brazil to Paris, France on June 1, 2009. The discussions that follow include a synopsis of each of these flights. These discussions examine the inaccurate situation assessments that the flight crews made during the critical phases of these flights. Recommendations are made for improving aircrew situation awareness training. Recommendations are also made for improving air crew alerting systems.

United air lines Flight 863, June 28, 1998

This narrative is based on information obtained from an article(Carley, 1999),(FAA ASR ACN: 406810, 1998), and from my experience flying United Air Lines(UAL) Boeing 747-400s out of San Francisco International Airport(SFO), (1992—1997). Some of the recommendations in this narrative are based on my experience as a flight instructor at the UAL Training Center, Denver, CO (1986-1992). UAL flight 863, weighing 850 thousand pounds, with a flight crew consisting of the Captain and three Captain qualified First Officers(FO), took off from SFO at 11:39 PM on June 28, 1987, on runway 28R. A FO in the right seat made the takeoff and initial climb. The other FOs served as observers. The weather at SFO was clear, but the hills west of the runway were obscured by a fog bank 1000 feet thick. The takeoff roll was normal. Just after liftoff and after landing gear retraction, the aircraft flew into zero visibility conditions and the crew heard a series of loud backfires. They also felt the aircraft severely vibrate. The crew first thought the noise and vibrations were caused by a blown tire, but after a short time, they determined that the number three engine had failed. The aircraft was at about 300 feet at that time. The Captain shut down the number three engine and the vibrations ceased. The Captain and one of the observer FOs started the engine shut down check list. At the same time, the other observer FO noticed that the airspeed was getting too low and told the flying FO to watch his speed. Shortly thereafter the stick shaker activated intermittently and the Ground Proximity Warning System (GPWS) sounded. The FO at the controls raised the nose of the aircraft and the GPWS ceased sounding. The Captain then took control of the aircraft and started to return the aircraft to its prescribed course and establish a safe climb speed. During the time that the crew dealt with shutting down the number three engine and restoring proper climb speed, the aircraft drifted off the prescribed route for engine failures. The aircraft had come within one hundred feet of hitting San Bruno Mountain, located fifteen miles northwest of SFO.

The critical phase of this flight started just after landing gear retraction and lasted about three minutes. The crew was distracted by backfire noises, severe vibrations, engine failure, and the low altitude shutdown of the failed engine. The crews' focus on these distractions caused them to lose track of the aircraft flight path. The FO at the controls was first to identify the engine failure, but he did not recognize the yaw condition caused by the engine failure and did not apply sufficient rudder input to control the turn away from the prescribed course. He tried to control the heading of the aircraft through the use of the aileron controls only. This raised the spoilers which increased drag and caused loss of airspeed. These distractions caused the crews' to make an inaccurate situation assessment concerning the critical task at hand; accurate control of the aircraft flight path. The incident described above revealed the necessity of re-emphasizing the importance of accurate flight path control in the face of compelling distractions. Crew training in flight simulators should include various engine failure situations at very low altitudes, in zero visibility conditions. Training in the use of proper rudder control in engine failure situations should also be included, as well as the use of the slip indicator in the Primary Flight Displays. Pilots should be thoroughly evaluated to insure they are highly proficient in handling a variety of engine failure scenarios. Flight crews that fly747-400s in international operations generally fly about four to five times per month and may not be familiar with one another. Emphasis should be placed on the importance of detailed pre-take off crew briefings concerning the duties of each crew member. These briefings should also include a detailed review of any special procedures for engine failures on departure.

Helios Airways Flight HCY 522, August 14, 2005

This narrative is based on findings from an accident report (Hellenic Republic Ministry of Transport & Communications Air Accidents Investigation and Aviation Safety Board [AAIASB] Helios Airways Flight HCY 522, 2006). Six crew members and 115 passengers were fatally injured in this accident. This narrative summarizes the critical phases of this accident and the situation awareness errors made by the flight crew. A description of factors that contributed to this accident, along with recommendations for modifying checklist procedures and cockpit warning indications are included.

Helios Airways Flight HCY522, a Boeing 737-300, with a crew of six and 115 passengers, took off from Larnaca, Cyprus at 0607UTC August 14, 2005. About seven minutes after takeoff, the aircraft was climbing through 16 thousand feet and the Cabin Altitude warning horn sounded.

The crew disengaged the Autopilot and Auto-throttle momentarily and then re-engaged them in modes to fly the programed flight route. As the aircraft continued climbing, the Captain contacted the airlines' operations base and reported a Takeoff Configuration Warning and a equipment cooling system problem. He requested the location of the circuit breakers for the equipment cooling system, and was informed that they were on a panel behind the captain's seat. As the aircraft climbed through 17 thousand feet, the Master Caution turned on, indicating that the passenger oxygen masks had deployed. The Master Caution was turned off, and the aircraft continued climbing. Communications with the operating base ceased as the aircraft climbed through 28 thousand feet, and the crew did not initiate or respond to any further communications for the rest of the flight. The aircraft continued to climb and leveled off at its programmed altitude of 34 thousand feet. The aircraft continued to fly the route programmed in the auto flight system for two hours and forty minutes, when the aircraft ran out of fuel and the aircraft crashed, killing all on board. Thirty minutes before the crash, Hellenic Air force F-16 fighters intercepted the aircraft, and observed the First officer slumped over the control wheel, and that the captain's seat was empty. They further observed that the passenger oxygen masks were deployed. The F-16 pilots tried to get the attention of the flight crew, but got no response. About fifteen minutes after the F-16s intercepted the aircraft the pilots observed a man who was not wearing an oxygen mask enter the cockpit and occupy the captain's seat. The man was believed to be the male flight attendant. The F-16 pilots tried without success to attract his attention. Five minutes before the aircraft crashed, the man started to respond to the F-16 pilots hand signals, but by then both of the aircraft's engines had flamed out, and 4 minutes later, the aircraft crashed.

This narrative has many examples of distractions that caused the flight crew to make an inaccurate assessment of the situation that resulted in this fatal accident. The Pilots reported for duty one hour before the flight's scheduled flight departure. This gave the crew only one hour to complete their preflight duties. These duties included flight- planning, a review of the aircraft maintenance logbook, preflight inspection of the aircraft, cockpit preparation for the flight and entry of the flight route into the aircraft flight management computer. The maintenance logbook contained an entry noting that the cabin pressurization system had been checked and tested for proper operation the evening preceding the flight. The accident investigation report (AAIASB, 2006, p. 116) concluded that the cabin pressurization Mode selector had been left in the MAN position by the maintenance crew. The accident investigation report (AAIASB, 2006, p. 117) also concluded that the flight crew failed to place the cabin pressurization Mode selector to the AUTO position during their prefight cockpit inspection. The report concluded that the cabin pressurization Mode selector remained in the MAN position until the aircraft crashed.

The flight crew's failure to detect the proper setting of the cabin pressurization system was the situation assessment error that distracted the crew from making accurate assessments of the critical events that led to their loss of control of the aircraft. The time for the crew's cockpit preflight inspection was limited, and they failed to notice that cabin pressurization mode selector was in the MAN position. They also failed to note the position of the mode selector when they performed the After Takeoff check list. After takeoff, at 12 thousand feet during the climb, the Cabin Pressurization warning horn sounded. The crew mistook the Cabin Pressurization warning horn for a Takeoff Configuration warning horn. While they were trying to deal with what they perceived as an aircraft configuration problem, they failed to respond to the indication that the cabin oxygen masks had deployed. They also failed to notice symptoms of hypoxia as the cabin altitude climbed to 28 thousand feet. The crew became incapacitated and remained so until the termination of the flight. The cabin crew did not establish communications with the flight crew immediately after the oxygen masks deployed. The confusion of the crew about the function of the Cabin Pressurization warning horn was instrumental in causing this accident. The accident report (AAIAS, 2006, pp. 104 – 110) provides several reports about crew confusion concerning the functions of the Cabin Pressurization warning horn and the Takeoff Configuration warning horn. The warning horn sounds intermittently for both conditions. While on the ground, the intermittent horn sounds only if the aircraft is not properly configured for takeoff. While airborne, the horn sounds only for a cabin pressurization failure.

I recommend that crews who fly versions of the Boeing 737-300 without an Engine Indicating and Crew Alerting System (EICAS) under take a thorough systems review of the 737-300 warning horn functions. I also recommend that a check for the proper position of the cabin pressurization mode selector be added to the Boeing 737-300 After Takeoff Checklist. I further recommend that crews be exposed to various subtle pressurization problems (e.g. gradual loss of cabin pressure) during training in flight simulators as well as the sudden loss of cabin pressure scenario. Annual training for all aircrews should include a review of hypoxia symptoms. In addition, flight attendants should be briefed to inform the pilots as soon as possible of a passenger oxygen mask deployment.

Air France Flight AF 447 June 1, 2009

This narrative is based on information from the final accident report (BEA, July 3, 2012) concerning the fatal crash of Air France flight AF 447. AF flight AF 447, an Airbus A330-203, with a crew of a Captain , two first officers (FO), nine flight attendants , and 216 passengers, departed Rio de Janeiro on May 31, 2009, bound for Paris, France. The aircraft climbed to 35 thousand feet and proceeded without incident for two hours and ten minutes. The Pitot probes then became obstructed by ice, causing inaccurate airspeed indications and disconnection of the Auto Flight System. The two FOs were flying the aircraft. They lost control of the aircraft and it crashed into the Atlantic Ocean four minutes and 23 seconds later, killing all on board.
The accident report (BEA pp. 22-24) provides a detailed narrative of the events that occurred from the time the Auto Flight system disconnected until the time the aircraft crashed. The following synopsis is a summary of this narrative.

After the Auto Flight system disconnected, the inaccurate airspeed indications continued, and the pilot flying (PF) for reasons unknown, put the aircraft into a steep climb to 38 thousand feet. During the climb, the PF trimmed the movable horizontal stabilizer to its maximum nose up position of 13 degrees, where it remained until the flight crashed. The stall warning sounded intermittently, and the aircraft entered a high angle of attack attitude and a low airspeed condition outside of the aircraft flight envelope This silenced the stall warning. At 38 thousand feet, the aircraft started to descend at about 10 thousand feet per minute and a forward speed of 110kts. The PF lowered the aircraft nose momentarily to an angle of attack within the aircraft flight envelope. The stall warning sounded briefly and ceased when the PF raised the nose back to the high angle of attack attitude. The aircraft remained in this attitude until impact with ocean.

The pilots were faced with numerous distractions (BEA pp.173-182) that caused them to inaccurately assess the situation they encountered. When the Auto Flight warnings and cautions sounded, the aircraft was in a stable pitch, roll and airspeed condition. Very little manual control input would have been required to keep the aircraft on a stable flight path. In my opinion, the pilots were distracted by the sudden necessity to manually control the aircraft, the numerous warnings and cautions presented along with unreliable airspeed indications. In the face of all these distractions, the pilots failed to immediately assess the implications of various airspeed indication failures or recognize the requirement to perform the Unreliable Airspeed Indication checklist. In the face of all this confusion, the PF put the aircraft into a steep climb which caused more distractions, leading to the fatal outcome of the flight. The Airbus 320s and 340s have histories of problems with Pitot probes (BEA p.124,p-.p.144). More attention should be addressed to improving these probes and the crew warnings associated these probes. These unreliable airspeed indication problems should be explicitly identified on the ECAM, as well the as the procedures required to deal with them. Above all pilots should be continually trained and reminded to make flight path control their first priority in any emergency situation

CONCLUSIONS

The incident and accidents discussed in these papers illustrate how difficult accurate situation assessment can be during emergencies, and how difficult it is to focus on basic flight path control when faced with multiple distractions. Research and training to promote accurate situation assessment in all phases of flight operation should be expanded and continued.

References

Nasa Asrs Data Base Online. (1989). ACN: 606810 Retrieved from http://www.asrs.nasa.gov/search/database.html.

Bureau dù Enquetes et d' Analyses pour la securite l' aviation civile Report on the accident on 1st June 2009 to the AirBus A330-203 registered F-GZCP operated by Air France flight AF 447 from Rio Janeiro-Paris. July 2012 Retrieved from http://www.bea.aero/en/enquetes/flight.af.447/rapport.final.en.php

Carley, W. M. (1999 March 19) . United 747's Near Miss Initiates A Widespread Review of Pilot Skills. The Wall Sstreet Journal, pp. 1-7.

Durso, F. T., & Sethumadhava, A. (2008). Situation awareness: understanding dynamic environments. Human Factors, 50(3), 442-448. Doi: 10.1518/001872008X288448

Hellenic Government Ministry of Transport and Communications Air Accident Investigation & Aviation Safety Board(AAIASB) Aircraft accident report Helios Airways flight HCY522 Boeing 737-318 at Grammatiko, Hellas on 14 August 2005. November 2006. Retrieved from http://www.moi.gov.cy/moi/pio/.../FINAL%20REPORT%205B-DB/.pdf

THE IMPACT OF BINAURAL BEAT TECHNOLOGY ON VIGILANCE TASK PERFORMANCE, MENTAL WORKLOAD, AND PSYCHOLOGICAL STRESS

Elizabeth A. Shoda
Wright State University
Dayton, OH

Gary N. Burns
Wright State University
Dayton, OH

Vigilance research dates back to WWII when psychologists attempted to explain why sonar radar operators were missing signals allowing German U-boats to pass undetected. This error in vigilance was termed the vigilance decrement. Since WWII the vigilance decrement has been responsible for a number of military, commercial, and industrial accidents and deaths. One possible area of interest which may help to solve this problem is called binaural beat technology. Binaural beats operate by entraining the brain in a frequency following response. Depending on the frequency of the binaural beat, different psychological and physiological results can occur. The current study examined the effects that beta binaural beats had on vigilance task performance, mental workload, and psychological stress. Results indicated that, under certain situations, participants listening to binaural beats during the vigilance task had significantly increased vigilance performance but rated the task as more challenging than the control group.

Vigilance can be defined as the ability to sustain attention and respond appropriately to demands in the environment (Shaw, Matthews, Warm, Finomore, & Silverman, 2010). Until recently, psychologists tended to view vigilance tasks as mentally undemanding (Heilman, 1995). Researchers believed that activity in brain systems were suppressed by the monotonous and repetitious nature of vigilance tasks. New research examining task type, perceived mental workload, resource demand, and task-induced stress has challenged this view (Warm, Parasuraman, & Matthews, 2008). Warm, Dember, and Hancock (1996) conducted a series of experiments showing that vigilance tasks are very resource demanding. They found that the vigilance decrement is accompanied by a linear increase in overall workload. Studies using neuroimaging methods, such as TCD, PET, and fMRI, also support the assertion that vigilance tasks are mentally demanding (Parasuraman, Warm, & See, 1998; Warm et al., 2008; Warm & Parasuraman, 2007). In addition, Szlama, Warm, Matthews, Dember, and Weiler (2004) found that scores on the NASA-Task Load Index were increasing significantly throughout the course of vigilance tasks. Since the discovery and acceptance of vigilance as a mentally demanding activity, research has been better able to explain the causes of the vigilance decrement. However, the issue of how to reduce or eliminate the vigilance decrement still remains largely unsolved.

One area of research that may be able to help solve this problem is brainwave entrainment (BWE). This term refers to the use of rhythmic stimuli with the intention of producing a frequency-following response of brainwaves to match the frequency of the stimuli (Huang & Charyton, 2008). Binaural beats are a particular type of BWE that has often been used in psychological research. When two similar, but different, frequencies are presented separately to each ear, the difference between them is perceived as a binaural beat (Wahbeh, Calabrese, & Zwickey, 2007). For example, if 300 hz are presented to the left ear and 290 hz are presented to the right ear, the listener will perceive a 10 hz binaural beat. In this scenario, the listener will have increased brainwave activity at the 10 hz frequency. Depending on the frequency of the binaural beat, different psychological and physiological effects will occur. The most commonly studied frequency ranges are delta frequencies (1-4 Hz) which are associated with deep sleep; theta frequencies (4-8 Hz) which are associated with light sleep, creativity, and insight; and beta frequencies (13-30 Hz) which are associated with a thinking, focused state and increased arousal (Huang & Charyton, 2008; Rangaswamy et al., 2002).

Past studies have demonstrated the effects of binaural beats on different psychological variables (Atwater, 2009, Wahbeh et al., 2007, Huang & Charyton, 2008); however, only one published study has ever examined the effects of binaural beats on vigilance performance (Lane, Kasian, Owens, & Marsh, 1998). This study found marginal support for the use of binaural beats to improve vigilance task performance; however, the study also posed several methodological concerns and raised questions as to why no one has replicated this study or conducted any other vigilance/binaural beat studies since. The current study aims to provide further support for the use of binaural beats to improve vigilance performance and open up new directions for future research.

Purpose and Hypotheses

The purpose of this study was to examine whether binaural beats have significant effects on vigilance task performance, psychological stress, and workload. This study replicated and expanded upon Lane et al. (1998) by using a between subjects design and by examining the effects of this technology on perceived workload and stress. It was hypothesized that participants listening to beta binaural beats would have better scores on the vigilance task as measured by hit rates and false alarms, as well as having lower scores on measures of psychological stress and mental workload when compared to the control group. The binaural beat and control conditions were further separated by a vigilance task that was either easier or more difficult based on event rate. The purpose of this was to determine whether event rate moderated the relationships between audio condition, perceived workload and stress, and vigilance performance.

Method

Participants

One hundred and thirty individuals participated in the experiment; 38.5% of participants were male, 50% were female, and 11.5% did not respond. The mean age for participants in this study was 20.87. All participants were enrolled at a large, public university in Southwest Ohio and completed the experiment for partial course credit. Participants were required to have normal or corrected-to-normal hearing and vision. These inclusion criteria ensured that participants could see and hear the stimuli presented. The online study sign-up included a section that explained these requirements to the participants. They were not able to participate in the study if the hearing and vision criteria were not met.

Design

This experiment utilized a 2x2x6 mixed design with two between subjects independent variables (audio condition, event rate) and one repeated measures independent variable (period) to examine vigilance performance across time and between conditions. Audio condition had two levels: a control condition with pink noise coupled with no binaural beats and an experimental condition with pink noise coupled with beta binaural beats. Event rate also had two levels: one fast and one slow event rate task. Period was a within subjects factor with each subject completing six continuous 5-minute trials. Participants were randomly assigned into one of four conditions: binaural beat, slow event; binaural beat, fast event; pink noise, slow event; pink noise, fast event.

Stimuli

Audio. The two audio tracks used in this study were created using MATLAB (The MathWorks Inc., 2010). Following procedures used in similar studies (Lane et al., 1998; Wahbeh, Calabrese, & Zwickey, 2007), the tracks contained a background of pink noise to mask the binaural beats. Binaural beats of 16 and 24 hz were used in this study for the binaural beat track. The tracks were played using stereo headphones and the intensity was set to a comfortable level, as determined by the participant. All audio stimuli and administration procedures were consistent with stimuli and procedures uses in Lane et al. (1998).

Vigilance Task. The vigilance task was created using MATLAB (The MathWorks Inc., 2010). The task involved participants watching the computer monitor and reacting when a target was present. Replicating Lane et al. (1998), individual capital letters were presented on the screen from a list of twenty six. The target occurred whenever a letter was repeated. For example, if the letters "A, R, G, M, M" were presented, the second "M" would be the target. The experimental vigilance task included six periods lasting five minutes each for a total of 30 minutes. In addition, the two different event rates used in this task were a quick 75 events/minute event rate representing the hard task and a slower 20 events/minute event rate representing an easier task. Both tasks used a 100ms display rate and 6% critical signal rate with the rule that two critical signals could not occur back to back. Participants were instructed to respond as quickly as possible to a critical signal.

Vigilance Performance Data

Vigilance performance was measured as the proportion of correct responses to critical signals (hits) and incorrect responses to noise signals (false alarms). Hits were determined by whether a participant responded to the critical signal within the allowed time (800ms).Responses in the absence of critical signals were coded as false alarms. Each participant had 800ms to respond to an event, regardless of which condition they were in.

Psychological Outcome Measures

Psychological stress was measured before and after the vigilance task using the Short Stress State Questionnaire (SSSQ; Helton, 2004). The SSSQ is a 24-item multidimensional questionnaire based on the Dundee Stress State Questionnaire (DSSQ; Matthews et al., 1999). The SSSQ factors include Distress, Engagement, and Worry. They are meant to closely reflect the concept of a mental trilogy composed of affect, conation, and cognition. Stress was also measured before and after the vigilance task using the Stressor Appraisal Scale (SAS; Schneider, 2008). The SAS is a ten item scale representing primary and secondary stressor appraisals. Primary appraisals are evaluations of how personally significant and relevant the situation is and items ask about situational threat, demand, stressfulness, exertion, effort, importance, and uncertainty. Secondary appraisals are evaluations of the amount of resources one has to cope with the situation and measures manageability, ability, and performance. Mental workload was measured after the vigilance task using a computerized version of the NASA Task Load Index (TLX; Hart & Staveland, 1988). The TLX is a multi-dimensional rating procedure that provides an overall workload score based on an average of ratings on six subscales: Mental Demand, Physical Demand, Temporal Demand, Own Performance, Effort, and Frustration.

Procedures

Participants were told that they were testing a new computerized vigilance task to assess its usefulness. They were not informed about hearing binaural beats and were told that the purpose of the audio track and headphones was simply to block out background noise. Before starting the vigilance task, participants completed pre-measures of the SSSQ and SAS. Participants then completed a short practice vigilance task. This practice task was 5 minutes in length and participants were trained in this short trial until they had at least an 80% hit rate and, at most, a 6% false alarm rate. Each participant was assigned at random to one of four treatment conditions. Participants either listened to an audio track containing only pink noise or a track containing pink noise with beta binaural beats. In addition, participants either completed the slow or fast event rate version of the vigilance task. The experimental vigil was 30 minutes long and involved participants sitting at a computer, pressing the spacebar as quickly as possible when the target was presented. Once the vigilance task was completed, the workload measure was presented followed by a second administration of the stress measures.

Results

Manipulation Check

Of the 130 participants in this study, data from 82 were analyzed, while 48 were excluded. In lieu of not having access to EEG technology, the participants were asked at the end of the experiment if they heard a "wobbly" noise or "beat" in their headphones. This "wobbly" noise was indicative of the presence of binaural beats. Participants in the binaural beat condition who did *not* report hearing this noise were excluded from analyses. In addition, data from participants in the pink noise condition were randomly selected to be included in the analyses to achieve equal n's within conditions. Using these selection criteria, near equal n's were obtained (PN, Slow: $n = 20$; BB, slow: $n = 20$; PN, fast: $n = 21$; BB, fast: $n = 21$).

Vigilance Performance

A 2 (audio condition) x 2 (event rate) x 6 (period) mixed-ANOVA was performed on the arcsines of the percentages of correct detections. For these, and all subsequent analyses, Box's Epsilon was used to correct for violations of sphericity. There were no significant main effects for audio condition, event rate, or their interaction. A significant main effect was found for period ($F(3.71,289.15) = 52.84$, $p < .001$), indicating that a vigilance decrement was present in this study. Significant interactions were found for period by audio condition ($F(3.71, 289.15) = 3.05$, $p < .05$), period by event rate ($F(3.71, 289.15) = 2.95$, $p < .05$), and period by audio condition by event rate ($F(3.71, 289.15) = 2.60$, $p < .05$). The 3-way interaction can be seen in Figure 1.

Figure 1.
Period by audio condition by event rate interaction for correct detections

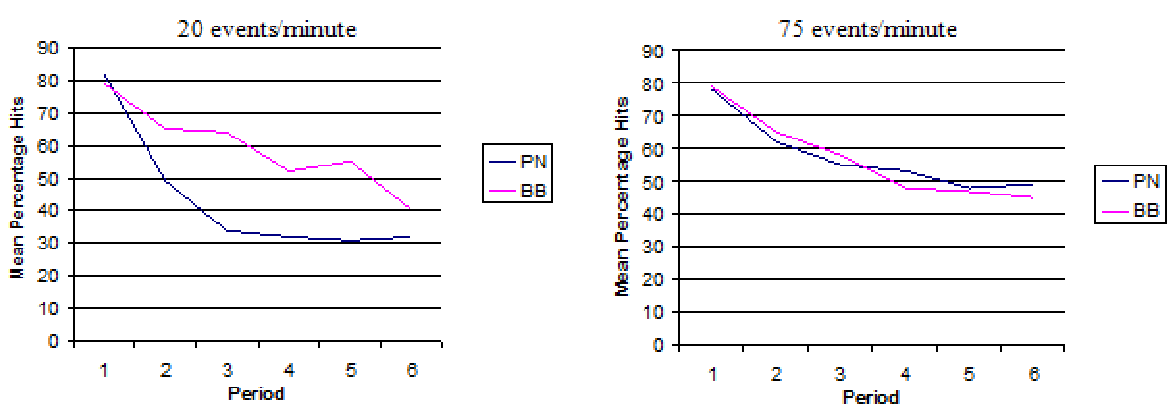


Figure 1 illustrates that participants listening to binaural beats performed somewhat better on the vigilance task than those listening only to pink noise, but for the slow event rate condition only. A series of follow-up t-tests indicated significant differences in the hypothesized direction during period two ($t(38) = 1.80$, $p < .05$), three ($t(38) = 2.65$, $p < .05$), four ($t(38) = 1.86$, $p < .05$), and five ($t(38) = 2.42$, $p < .05$). No significant effects were found for false alarms.

Psychological Outcome Variables

For the SSSQ, 2 (audio condition) x 2 (event rate) between-subjects ANOVAs were performed on the pre test scores of each subscale. No pre-test differences existed between groups. To test whether the experimental manipulations had any effect on psychological stress, difference scores were computed by subtracting pre-manipulation scores from post-manipulation scores. A 2 (audio condition) x 2 (event rate) between-subjects ANOVA was then conducted on the difference scores of each subscale. For Engagement, the pre-post change significantly varied by event rate ($F(1,78) = 10.17$, $p < .01$) with participants in the slow event rate ($M = -.41$, $SD = .50$) showing stronger decreases than participants in the fast event rate ($M = -.09$, $SD = .40$). There were no significant differences for Distress or Worry. Similarly, 2 (audio condition) x 2 (event rate) between-subjects ANOVAs were conducted on the pre-test scores of the SAS subscales and revealed no significant differences. For post-pre difference scores, 2 (audio condition) x 2 (event rate) between-subjects ANOVAs revealed that, for primary appraisals, there were main effects for both audio condition ($F(1,78)=5.22$, $p<.05$) and event rate ($F(1,78)=4.28$, $p< .05$), but not their interaction. Results indicate that mean difference scores were higher in the binaural beat condition ($M = 1.03$, $SD = .96$) than the pink noise condition ($M = .48$, $SD = 1.17$) and in the fast event rate condition ($M = 1.0$, $SD = 1.1$) than the slow event rate condition ($M = .51$, $SD = 1.1$). These difference scores indicate that participants in the binaural beat and fast event rate conditions thought that the completed task was more challenging than they originally expected. No significant effects were observed for secondary appraisals.

For the NASA-TLX, a 2 (audio condition) x 2 (event rate) x 6 (subscale) mixed-ANOVA revealed a significant main effect for subscale ($F(4.07, 317.58) = 20.95$, $p<.001$) and a significant subscale by event rate interaction ($F(4.07, 317.58) = 2.41$, $p<.05$). Subsequent analyses revealed that the subscales contributing to the significant factor by event rate interaction were Effort ($t(80) = -2.19$, $p<.05$) and Temporal Demand ($t(80) = -3.05$, $p<.01$). Scores for Temporal Demand were higher in the fast event rate condition ($M = 57.53$) than the slow event rate condition ($M = 39.33$), and scores for Effort were higher in the fast event rate condition ($M = 59.70$) than the slow event rate condition ($M = 43.9$). No significant effects were observed for audio condition.

Discussion

This study sought to investigate the effects of binaural beat audio technology on vigilance task performance, psychological stress, and mental workload. This study also sought to replicate the vigilance performance results of Lane et al. (1998). Hypothesis 1 stated that individuals listening to beta binaural beats would have improved vigilance task performance compared to individuals listening only to pink noise. Results of this study partially confirmed this hypothesis. Although no significant main effect differences existed for hit rate, a significant factor by audio condition interaction revealed hit rate differences as time on task continued. Participants in the binaural beat condition experienced a delayed vigilance decrement compared to participants in the pink noise condition, and this effect was moderated by event rate. In contrast, Lane et al. (1998) found significant main effects for binaural beats, but no interaction effects. Although both studies demonstrated significant performance increases for participants listening to binaural beats, the mechanism for increased performance varied. Hypothesis 2 stated that participants listening to binaural beats would have decreased psychological stress and workload after the task when compared to those listening to pink noise. This hypothesis was not confirmed.

The results of this study indicate that during a slow event rate task, binaural beats have the ability to reduce the negative effects that the vigilance decrement has on performance. When a fast event rate is used, this effect disappears. Although these initial results are promising, more research needs to be conducted to validate and expand on this study. Specifically, subsequent experiments should be conducted which examine if these effects can be replicated with shorter or longer vigils, different binaural beat frequencies, and different types of vigilance tasks. Ultimately, if these effects can be consistently demonstrated, the deleterious effects of the vigilance decrement on workplace performance may be reduced.

References

Atwater, H. F. (2009). Binaural beats and the regulation of arousal levels. *TMI Journal*, 1, 1-17.

Hart, S. G., & Staveland, L. E. (1988). *Development of NASA (Task Load index): Results of empirical and theoretical research.* In P. A. Hancock & N. Meshkati (Eds.), Human Mental Workload (pp. 239-250). Amsterdam: North Holland Press.

Heilman, K. M. (1995). Attentional asymmetries. In R. J. Davidson & K. Hugdahl (Eds.), *Brain asymmetry* (pp. 217-234). Cambridge, MA: MIT Press.

Helton, W. S. (2004). Validation of a short stress questionnaire. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, 1238-1242.

Huang, T. L, & Charyton, C. (2008). A comprehensive review of the psychological effects of brainwave entertainment. *Alternative Therapies in Health and Medicine*, 14, 38-50.

Lane, J. D., Kasian, S. J., Owens, J. E., & Marsh, G. R. (1998). Binaural auditory beats affect vigilance performance and mood. *Physiology & Behavior*, 63, 249-252.

Matthews, G., Joyner, L., Gilliland, K., Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a state big three? In I. Merville, I.J. Deary, F. DeFruyt, and F. Ostendorf (Eds.), *Personality psychology in Europe* (vol. 7) pp. 335-350. Tilburg: Tilburg University Press.

Parasuraman, R., Warm, J. S., & See, J. E. (1998). Brain systems of vigilance. In R. Parasuraman (Ed.), *The attentive brain* (pp. 221-256). Cambridge, MA: MIT Press.

Rangaswamy M, Porjesz B, Chorlian DB, Wang K, Jones KA, Bauer LO, Rohrbaugh J, O'Connor SJ, Kuperman S, Reich T, Begleiter (2002). "Beta power in the EEG of alcoholics". *Biological Psychology,* 52, 831–842.

Schneider, T. R. (2008). Evaluations of stressful transactions: What's in an appraisal? *Stress and Health,* 24, 151-158.

Shaw, T. H., Matthews, G, Warm, J. S., Finomore, V. S., & Silverman, L. (2010). Individual differences in vigilance: personality, ability, and states of stress. *Journal of Research in Personality*, 44, 297-308.

Szalma, J. L., Warm, J. S., Matthews, G., Dember, W. N., & Weiler, E. M. (2004). Effects of sensory modality and task duration on performance, workload, and stress in sustained attention. *Human Factors*, 46, 219-233.

Wahbeh, H., Calabrese, C., & Zwickey, H. (2007a). Binaural beat technology in humans: a pilot study to assess psychological and physiological effects. *The Journal of Alternative and Complementary Medicine*, 13, 25-32.

Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automated and human performance: Theory and applications* (pp. 183-200). Mahwah, NJ: Erlbaum.

Warm, J. S., & Parasuraman, R. (2007). Cerebral hemodynamics and vigilance. In R. Parasuraman & M. Rizzo (Eds.). *Neuroergonomics: The brain at work* (pp. 146-158). New York: Oxford University Press.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors*, 50, 433-441.

ABLE FLIGHT: INCREASING AVIATION OPPORTUNITIES FOR PEOPLE WITH DISABILITIES

Jennifer Kirschner
Department of Aviation Technology, Purdue University
West Lafayette, IN, USA

Branden Avery
Department of Aviation Technology, Purdue University
West Lafayette, IN, USA

Wes Majors
Department of Aviation Technology, Purdue University
West Lafayette, IN, USA

Erin Bowen
Department of Technology Leadership and Innovation, Purdue University
West Lafayette, IN, USA

Bernard Wulle
Department of Aviation Technology, Purdue University
West Lafayette, IN, USA

Access to certain aviation careers has traditionally been limited for people with disabilities. Although the Americans with Disability (ADA) Act of 1990 mandated that disabled applicants be given equal consideration for jobs for which they can perform the essential job functions, stereotypes and lack of knowledge about career opportunities may prevent qualified disabled people from apply in the first place. In order to remove perceived barriers to participation and to increase the visibility of opportunities for both life success and employment within the aviation industry for people with disabilities, the Purdue University Aviation Technology department has begun to participate in the Able Flight program. Able Flight offers a select number of applicants with significant physical impairments each year scholarship funds toward aviation-related training. This longitudinal study will survey former Able Flight participants to estimate the effect that training has had on their career aspirations, either in aviation or other industries.

A large percentage of Americans live with some type of disability. Compared to their able-bodied counterparts, they may be more likely to be unemployed or underemployed, due to a number of factors. Employer perceptions and lack of accessible facilities may limit career choices. Even though federal changes to employment laws have increased access to employment opportunities, hidden biases may negatively impact the perception of disabled individuals during the selection process. Simply having the opportunity to make career-related choices and decisions, however, does not ensure that a person will have the self-determination to actively pursue those opportunities and overcome potentially negative perceptions. Educational interventions designed to increase autonomy, competence, and relatedness have the potential to provide transferable skills that increase self-confidence and the willingness to try again. Additionally, it may also help raise awareness of the need for better access to employment opportunities for disabled individuals.

Literature Review

A vast number of people in the United States live with a disability. Over the years laws have been enacted to facilitate job opportunities for the disabled, but individuals with a disability still remain scarce in the employment sector. Statistics provided by the Census Bureau makes this apparent.

Disabilities are extremely prevalent within the United States. According to the 2011 American Community Survey (ACS), a yearly survey conducted by the Census Bureau, out of 241 million individuals age 16 and over currently living in the United States, almost 35 million are living with a disability (U.S. Department of Commerce, 2011). Thus, it is estimated one out of every 14 people age 16 and older presently deal with some sort of disability on a daily basis, making persons with a disability one of the largest minorities in the United States. The ACS lists six disability types - hearing, vision, cognitive, ambulatory, self-care, and independent living difficulty.

The Bureau of Labor Statistics reported 17.8% of persons with a disability were employed during 2011, less than the previous year's employment-population ratio of 18.6% (U.S. Department of Labor, 2011). In contrast, almost 64% of non-disabled persons were employed during this time period. In 2011 the unemployment rate for individuals with a disability was almost double the unemployment rate for their able bodied counter parts (8.7% versus 15.0%). These rates remained unchanged from the previous year for individuals with a disability, while the rates for individuals without a disability dropped. Especially in turbulent economic times, unemployment rates among the disabled can rise sharply.

The industries with the highest levels of representation of disabled employees were the education and health services, retail services, as well as business and professional services; 21.9%, 12.8%, and 10.6% respectively. The types of occupations that were most likely to employ persons with a disability were - management, 31.7%, followed by sales and office related work, 24.6%. The majority worked in the private sector (72.5%), while 15.6% had government jobs, and 11.8% were self-employed (U.S. Department of Labor, 2011).

Individuals with a disability are underrepresented and underemployed in the job market, even with legislation to ensure equal opportunity hiring. This is made obvious when the Bureau of Labor Statistics released 2011's employment-population ratio for individuals with a disability. Additional evidence is found in the unemployment rates for the handicapped, which are nearly doubled that of non-disabled individuals. Although disabled individuals are willing to work, a large portion has not been given the opportunity to do so. Preconceived opinions and perceptions by employers may have left a potentially productive talent pool untapped.

Since the Rehabilitation Act of 1973, hiring decisions involving physically disabled applicants have been subject to much research. Studies have shown mixed results; some managers have shown positive reactions toward hiring disabled employees, and others are more negative.

Depending on the severity of the disability, those with disabilities can be limited in the activities that they can perform during a normal work day. In certain work environments, the ability to perform physically demanding or socially demanding tasks may be considered “essential” job functions, which limits their accessibility to those with disabilities. In these instances physically disabled persons may be limited to very specific offers of employment (Drehmer & Bordieri, 1985). Additionally, Florian (1978) conducted a survey that concluded that, regardless of the severity, the cause of a disability was directly related to workforce acceptance. Those suffering paralysis due to a war related injury were more likely to receive offers of employment than those with an external source injury (such as a car accident), even though the injuries were the same.

While some studies that requested hiring managers or business students to compare resumes of disabled applicants with non-disabled disabled ones have found a marked preference for disabled applicants (Krefting & Brief, 1977; Premeaux, 2001; Stone & Sawatzki, 1980), experimental research has found the opposite bias to be true during interviews (Johnson & Heal, 1976) Johnson and Heal (1976) examined different interviews that took place between different employers and an actor. The results from this study provided evidence that when the actor in a wheelchair was interviewed, he was viewed negatively and offered fewer employment opportunities.

There are several reasons for the different findings from the studies, mostly due to the difference in methods. A person who is wheelchair-bound may be perceived as less threatening on paper than in a face-to-face encounter. Block and Yuker (1979) explain that prejudice towards a person with disability is normally hidden, particularly in survey data, because it is not socially desirable or acceptable to display such bias.

Although federal amendments to employment law and changing societal norms have led to increases in the availability of employment opportunities for the physically disabled, simply having the opportunity to make career-related choices and decisions does not ensure that a person will have the self-determination to actively pursue those opportunities and overcome potentially negative perceptions, much the same way that having equal access to employment opportunities does not ensure that one will be able to find a suitable job. The concept of self-determination is a very common theme in the narrative of both opportunities and career success for those with disabilities. It can be defined as “acting as the primary causal agent in one’s life and making choices and decisions regarding one’s quality of life free from undue external influence or interference” (Wehmeyer, 1997, p. 177), and is an outgrowth of follow-up research conducted after the first generation of children completed schooling under the Education for All Handicapped Children Act in 1975. Now that access to educational opportunities was federally mandated, researchers hoped to find disabled individuals acting as productive, integrated members of society. While rates varied between states, longitudinal studies showed that the vast majority of youths with disabilities were either unemployed or underemployed, and most lived in residential communities, instead of living independently, even after finishing high school. Self-determination was shown to distinguish between those who held jobs and lived independently, and those who did not.

Self-determination theory (SDT) asserts that people are inherently drawn towards personal development and growth, but that this need must be nurtured in order for it to develop properly. Three basic drives or motivations that pull us toward personal growth are our need for competence, relatedness, and autonomy (Deci & Vansteenkiste 2004). Competence is a need for task mastery. Relatedness is a universal urge to be connected in a meaningful way to those around us. Autonomy is not necessarily a need for independence, but a desire to be a change agent in one's own life. In order to meet these three needs, there are two different types of motivation: intrinsic and extrinsic (Perreault & Vallerand, 2007). Intrinsic motivation is an internal drive to complete tasks associated with cognitive and social development, because they are inherently enjoyable. In order for tasks to be perceived as being caused by intrinsic motivation, both competence and autonomy must be present. Extrinsic motivation, on the other hand, refers to external pressure to complete a task, either from parents, peers, or others. Extrinsic motivation can be accompanied by differing levels of autonomy, competence, and relatedness, but in order for goals to be internalized as meaningful, all three should be high.

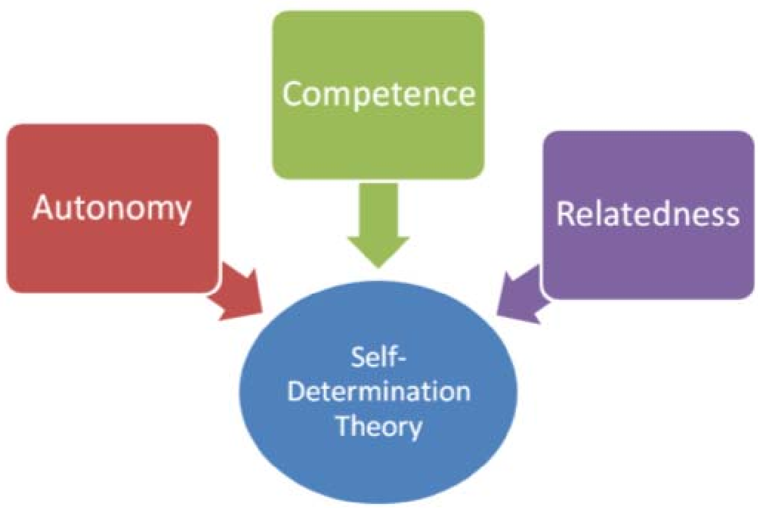


Figure 1. Self-determination theory.

Opportunities to practice self-determination skills can increase self-confidence and the willingness to try again (Stoner, Angell, House, & Goins, 2009). Educational interventions can also increase self-determination, and should be designed to foster developing decision-making skills, goal-setting skills, risk-taking skills, self-observation and evaluation, and an internal locus of control (Wehmeyer, 1997).

People with disabilities have been recognized as productive employees in many other industries, but have not yet been accepted into aviation careers in large numbers. While deaf pilots have been able to gain some traction within the industry (Moore, 2012), people with other disabilities can have a much more difficult time attempting entry. The Federal Aviation

Administration recently announced that it has set a goal of hiring at least three percent per fiscal year for individuals with severe disabilities (Federal Aviation Administration, 2012).

Proposed Methodology

In order to more clearly examine the effect that a focused skills-based training intervention has on the career aspirations and self-determination of those with disabilities, this research study will survey participants who have received either a flight or a career training scholarship from Able Flight, a non-profit organization that seeks to increase the accessibility of aviation. Since 2006, approximately 20 students with significant physical impairments have received training scholarships toward attaining a Sport Pilots licenses, maintenance certificates, and dispatcher licenses. Able Flight's "mission is to offer people with disabilities a unique way to challenge themselves through flight training, and by doing so, to gain greater self-confidence and self-reliance" (Able Flight, 2011, para. 1).

This longitudinal study will seek to determine the effect that such training had made on the career paths that students choose, either in aviation or in another industry. Survey participants will be contacted using email, and surveys will be distributed using Qualtrics, in order to increase their accessibility. The survey will use both open ended questions and Likert scale responses. Participants will be asked to discuss the specific skills they gained during the training intervention and what career paths or opportunities they have taken since completion. Other areas covered include what aviation careers they see as being accessible to those with disabilities, how the training intervention impacted their view of themselves, and, if they are not currently employed in aviation, what skills gained during the training intervention were transferable to their current employment situation.

Survey data, once gathered, will be analyzed using the NVivo qualitative data analysis software package. Particular emphasis will be placed on mentions of the three drives in self-determination theory (autonomy, competence, and relatedness) to determine if these three areas were important to the Able Flight participants, and if the training intervention led to changes in the career opportunities that they see as being accessible, even if they have not been able to pursue a career yet.

Conclusions

In conclusion, when completed, this research project will seek to better understand the employment opportunities available for and perceptions of those with disabilities. Additionally, it will attempt to better define the means by which educational interventions such as Able Flight impact the self-determination and perceived career opportunities available to those with disabilities. Research of this nature has the potential to better enable educators to design future interventions that facilitate self-determination in those with disabilities. Lastly, it may also help raise awareness of the need for better access to employment opportunities for disabled individuals.

References

Able Flight. (2011). *Able Flight: About Us.* Retrieved from http://ableflight.org/about-us

Bordier, J. E., & Drehmer, D. E. (1986). Hiring decisions for disabled workers: Looking at the cause. *Journal of Applied Social Psychology, 16*(3), 197-208.

Deci, E. L., & Vansteenkiste, M. (2004). Self-determination theory and basic need satisfaction: Understanding human development in positive psychology. *Ricerche di Psichologia, 27,* 17–34.

Drehmer, D. E., & Bordieri, J. E. (1985). Hiring decisions for disabled workers: The hidden bias. *Rehabilitation Psychology, 30*(3), 157-164.

Federal Aviation Administration. (2012). *People with Disabilities Program*. Retrieved from https://www.faa.gov/about/office_org/headquarters_offices/acr/outreach/pwd/

Moore, J. (2012, January 25). *Deaf pilot spreads word: You can fly.* Retrieved from http://www.aopa.org/aircraft/articles/2012/120125deaf-pilot-spreads-the-word.html

Perreault, S. & Vallerand, R. J. (2007). A test of self-determination theory with wheelchair basketball players with and without a disability. *Adapted Physical Activity Quarterly, 24*, 305-316.

Premeaux, S. F. (2001). Impact of applicant disability on selection: The role of disability type, physical attractiveness, and proximity. *Journal of Business and Psychology, 16*(2), 291-298.

U.S. Department of Labor, Bureau of Labor Statistics. (2011). *Persons with a disability: Labor force characteristics – 2011* (USDL-12-1125). Retrieved from http://www.bls.gov/news.release/pdf/disabl.pdf

U.S. Department of Commerce, United States Census Bureau. (2011). *Selected economic characteristics for the civilian noninstitutionalized population by disability status* (S1811). Retrieved from http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_11_1YR_S1811&prodType=table

Stoner, Angell, House, & Goins. (2009). Self-determination: Hearing the voices of adults with physical disabilities. *Physical Disabilities: Education and Related Services, 25*(1), 3-35.

Wehmeyer, M. (1997). Self-determination as an educational outcome: A definitional framework and implications for intervention. *Journal of Development and Physical Disabilities, 9*(13), 1997.

IMPROVING STUDENT RETENTION WITH THE CREATION OF A STUDENT CHAPTER OF A PROFESSIONAL ORGANIZATION

John H. Mott
Department of Aviation Technology, Purdue University
West Lafayette, IN

Hunter R. Musselman
Department of Aviation Technology, Purdue University
West Lafayette, IN

Significant research demonstrating the connection between student retention and student-peer interaction is extant. Such interaction is facilitated by the creation and growth of a student chapter of a professional organization that welcomes students who might not be otherwise engaged on campus and encourages their involvement in the organization. Benefits are multiplied when the organization is located at a joint facility which houses a lower two-year program feeding an upper two-year completion program, as it can be shown that the students groups who benefit the most from student-peer interaction are students indigenous to the institution and transfer students from two-year institutions. We will discuss how a student chapter of the American Association of Airport Executives was created at our facility, and the benefits in terms of both student retention and opportunities for career advancement that have resulted at both the two-year and four-year institutions that are located therein.

According to data from ACT, Inc., national first- to second-year retention rates among students enrolled in public universities in the U.S. average 55.4% for two-year institutions and 65.6% for four-year institutions (ACT, 2011). In addition, approximately half of all undergraduate students who enter public universities fail to graduate within six years. While many colleges and universities have instituted intervention programs in an effort to improve retention rates, such programs have met with only marginal success (Seidman, 2005). Therefore, it is clear that both academic institutions and the students they serve will benefit from a coordinated effort to improve these retention rates.

Identification of an Appropriate Engagement Mechanism

There has been a considerable effort by educational researchers over the past twenty-five years to understand the factors that influence student success in academic programs and to finds means of positively influencing those factors. Astin (1984) has proposed a student development theory based on student involvement, defined as “the quantity and quality of the physical and psychological energy that students invest in the college experience,” and suggests that “the greater a student’s involvement in college, the greater will be the amount of student learning and personal development.” According to Astin, the student involvement theory explains most of the empirical data assimilated by researchers about environmental influences on student development. An important corollary to this theory, then, is that the educational policies that are most effective are those that are directed toward increasing student involvement.

Kinzie (2010) asserts that a positive correlation exists between student engagement and student persistence from the first to the second year of college, the transition point in the undergraduate experience where the attrition rate is typically the highest (ACT, 2011). This assertion is supported by Kuh, et al (2007), who also indicate that, “while exposure to educationally effective practices is associated with desired outcomes for all students, historically underserved students benefit more from engaging in these activities than white students in terms of earning higher grades and persisting to the second year of college.” It is therefore perhaps even more critical to ensure that effective policies that allow and encourage the involvement of underserved students in campus programs be implemented.

Bennett, et al. (2011) determined that 30% of students surveyed in the 2011 National Survey of Student Engagement conducted by the Indiana University Center for Postsecondary research spent at least 6 hours per week participating in co-curricular activities, including activities of campus organizations.

Chickering and Gamson (1987) provide seven principles that comprise good practices in higher education. Among these principles are the encouragement of student-faculty contact, the encouragement of cooperation among students, and the respecting of diverse talents. Such principles may be considered inherent in the creation and the continued development of student chapters of professional organizations, as these chapters require the involvement of a faculty advisor, thereby facilitating student-faculty contact, and require interaction and cooperation between the student members themselves. Therefore, it stands to reason that a student chapter of a professional organization fulfills several of the attributes suggested as good higher education practices by Chickering and Gamson.

Program Background

Roberts and McNeese (2010) showed that educational origin influences the degree of involvement of students in programs. According to that study, students indigenous to programs tend to be most involved in those programs, while students who have transferred from two-year community colleges tend to have the next highest level of involvement. Transfer students from four-year colleges tend to be the least involved of the student groups.

Economic realities are causing some educational institutions to reevaluate their provision of all four years of a traditional Bachelor's degree program in Technology. In locations where such programs are in close collaboration with two-year partner institutions, it makes sense, based on the research of Roberts and McNeese (2010) described previously, to facilitate a much higher level of curriculum integration between programs than has previously occurred. The Aviation Technology Department at Purdue University has successfully made recent efforts to strengthen its relationship with its partner institution in Indianapolis, Vincennes University. The aviation technology programs of both programs in Indianapolis are housed in a single facility located at the Indianapolis International Airport, the Aviation Technology Center.

Students matriculating in the Vincennes Indianapolis program may pursue Aviation Maintenance, Professional Flight, or General Studies options. Upon completion of an Associate's degree, these students can apply for admission into the Purdue Aviation Technology program, where they can finish the two years remaining toward their Bachelor's degree in Aviation Technology. Processes and systems at the Indianapolis facility shared by Purdue and Vincennes are being redeveloped to provide as seamless a transition between the two programs as possible. This is planned to include the future articulation of courses between programs.

The Purdue portion of the Aviation Technology Center offers the Aviation Operations Technology (AOT) major. This major is a plus-two management-focused degree program designed specifically to serve as a convenient opportunity for degree completion for students and working professionals who have not attained their desired educational goals. The program is intended to provide leadership education at the baccalaureate level to students with technical backgrounds in aviation.

Graduates in Aviation Operations Technology are responsible for the management of many different types of financial, capital, human, and information resources in the aerospace industry. They manage engineers and technologists who design, test, and build new aircraft and components. They may be responsible for managing the support of production aircraft, an area that includes product support, modification, and accident investigation. They may also serve as managers at airlines, airports, or general aviation operations around the world.

Matriculation channels, which are sufficiently diverse to ensure a higher level of opportunity for those students with prior college credit who wish to continue their college education, include the acceptance of students with associate degrees in aviation and in other fields, as well as the acceptance of transfer students with appropriate transfer credit or equivalent experiential credit. The AOT program consists of 124 credit hours, approximately 60 hours of which is delivered in the AOT program, with the remaining 64 hours generally considered to be either credit from other institutions, experiential credit, or both.

Courses leading to the completion of this degree are delivered through a variety of means, including traditional, hybrid, and distance (both synchronous and asynchronous) delivery methods, and are made available to traditional aviation technology students in Indianapolis and West Lafayette, and to distance students in many other locations.

Development of a Student Professional Organization Chapter

At the Aviation Technology Center, the commuter college atmosphere and short-term tenure of the students associated with each of the collocated state universities' two-year degree programs resulted in limited opportunities for student engagement. Because this engagement is an integral piece of many college students' education, the need for an all-inclusive student organization to provide such opportunities outside of the classroom was perceived by both administrators and students in the Aviation Operations Technology program. Shortly thereafter, in the summer months of 2010, a group of students interested in helping to improve opportunities for student engagement and guided by location administrators organized its ideas and broke new ground by successfully applying for and establishing the Aviation Technology Center's first student organization, a student chapter of the American Association of Airport Executives (AAAE). This organization has continued to grow since it was established in October of 2010.

In launching the organization, many resources were needed; among them were a student membership, officers, a faculty advisor, organizational structure, and chapter bylaws. A significant amount of student initiative and effort was exerted in the successful establishment of the organization. In the subsequent weeks, a membership call-out, chapter meetings and officer elections were held. At these initial meetings, the organization's executive committee effectively instituted a professional atmosphere, using standard parliamentary procedure to maintain order. The active participation of student members, involvement and guidance provided by the faculty advisor, and high degree of professionalism present throughout the development phase were key factors in the organization's initial and continuing success.

Methods Used to Get Students Involved in the Chapter

One of the unique aspects of the student organization in this example is that it is jointly sponsored by both collocated institutions. Joint sponsorship leads to an increase in the inherent complexity of the organization, and this translates to recruiting and membership functions. However, the most successful methods of student membership recruitment the organization was able to implement were also the simplest. These included exhibit tables at Vincennes University and Purdue University Open House events, student volunteers working at local air shows, and other similar methods. The marketing techniques utilized were also relatively simple, but effective. For example, student members explained to prospects the nature of the AAAE professional organization and how joining it would specifically benefit the prospects. Some of those benefits include both professional and community volunteering opportunities, development of their personal networks, and improvement of their individual professional skills, including communication, project development, professionalism, and more.

Various other methods of membership recruitment and retention have been utilized over several semesters. This student chapter is currently one of the few AAAE student chapters to utilize social networking, and manages both Facebook and Twitter pages. The organization delegates responsibilities of maintaining the pages, running advertisements, and promoting events and tours to the chair of the publicity committee, one of four committees. This specific position requires planning and development of brochures, posters, pictures, and video in order to successfully reach the 170+ students that AAAE regularly contacts through email.

The AAAE student organization has employed a variety of methods to engage student members, including Internet-based tools such as ProBoards, a discussion forum, and SurveyMonkey, a survey tool, to pose questions to members and solicit feedback. However, the group's leaders have found that the most effective method for getting students truly involved and interested is face-to-face communication. Current members have given classroom PowerPoint presentations to many students at the Aviation Technology Center to improve organizational visibility and recruit new members. In addition, the student chapter has taken advantage of multiple opportunities to present to and network with high school students attending local area career centers in aviation-oriented programs at the junior and senior levels. In its participation with these experiences, the AAAE leadership have noted that virtually all of these high school students have requested inclusion on the member e-mail list and have contacted the organization wanting to be further involved. In addition, the Federal Aviation Administration (FAA) held the first annual ACE Academy, a week-long aviation summer camp intended for high school students with an interest in aviation, at the Aviation Technology Center in 2011. The AAAE student chapter was heavily involved in the camp, with members serving as student volunteers at the event to promote aviation and serve lunches. The event was a success, and led to many young adults signing up for the AAAE member e-mail list and "liking" the organization on Facebook. AAAE continues to maintain e-mail communication with these students as they move closer to making their college choices.

Although getting under-motivated college students, many of whom work part-time jobs, enroll in maximum credit hours per semester, and commute to campus, to become involved with a student professional organization and to make positive strides in improving their lives is difficult, the challenge has been aptly met by AAAE. The organization initially faced various difficulties one would expect with any new group, but with dedicated leadership by the executive committee, interested and committed members, an excellent faculty advisor and a welcoming aviation community in Indianapolis, AAAE has accomplished many of its goals; most importantly, it has achieved a significant level of student member involvement.

Measures of Student and Program Success

While it is difficult to quantitatively measure student and program success after only a year and a half of operation, one may empirically judge the relative success of the organization's members, with some portion of that success directly attributable to involvement in AAAE. For example, the first president of the chapter established connections with industry and received an internship with a major aviation MRO facility in Indianapolis; that internship subsequently developed into a full-time position upon the student's graduation. Similarly, another charter member, having been heavily involved with the student chapter in an executive capacity, was offered an internship with a general aviation company in which he excelled. Student matriculation from the lower two-year program into the upper has improved, as well, increasing from 5.5% of the completion program's composition at the initiation of the student organization to 20.9% in the most recent semester for which enrollment data is available.

There are numerous benefits of belonging to an organization of AAAE's stature. Such membership can clearly lead to opportunities not otherwise available in an individual's career. For example, half of the chapter's current executive members will be attending graduate school upon graduation. Those individuals might not have otherwise considered post-baccalaureate educational opportunities had it not been for their positive contributions and successes resulting from their membership in AAAE.

One of the initial goals of the student AAAE chapter was to establish opportunities for students to "get out of the classroom," allowing them to make contact with industry professionals and experience elements of the aviation industry that they might not otherwise be able to experience in a traditional classroom setting. These networking and industry experiences are critical to students' continuous growth as they enter careers in the aviation industry. With over 20 active members and additional involvement from 170 high school students, college students and college graduates, the AAAE student chapter at the Aviation Technology Center is well-positioned to function as an important tool for student engagement and success in the future.

References

ACT, Inc. (2011). National collegiate retention and persistence to degree rates. Retrieved from http://www.act.org/research/policymakers/pdf/retain_2011.pdf

Astin, A.W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel*, 25, 297–308.

Bennett, D., Ikenberry, S., Broad, M. C., Sapp, M., Ewell, P., Torney-Purta, J., & Howard, M. A. (2011). Fostering Student Engagement Campuswide. *Higher Education Management*, 1-50.

Chickering, A.W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin*, 39, 3–7.

Kuh, G. D., Kinzie, J., Cruce, T., Shoup, R., & Gonyea, R. M. (2006, July). Connecting the dots: Multi-faceted analyses of the relationships between student engagement results from the NSSE, and the institutional practices and conditions that foster student success: Final report prepared for Lumina Foundation for Education. Bloomington, IN: Indiana University, Center for Postsecondary Research.

Roberts, J., & McNeese, M. N. (2010). Student Involvement/Engagement in Higher Education Based on Student Origin. *Research in Higher Education Journal*, *7*(1), 1-12. Retrieved from http://www.aabri.com/manuscripts/09346.pdf

Seidman, A. (2005). *College student retention: Formula for student success*. (A. Seidman, Ed.) *American Council on Education Oryx Press Series on Higher Education* (p. 350). Greenwood Publishing Group. Retrieved from http://books.google.com/books?id=RGCI36TwZh0C&pgis=1

A COLLABORATIVE FORUM FOR THE DISTRIBUTION OF MULTIDISCIPLINARY SCHOLARSHIP IN THE OPEN-ACCESS ENVIRONMENT: THE INCEPTION OF THE ADVANCED AVIATION ANALYTICS INSTITUTE FOR RESEARCH (A^3IR-CORE)

Brent Bowen
Erin Bowen
Thomas Carney
Jennifer Kirschner
Chien-tsung Lu
Reilly Meehan
John Mott
Clay Wildt
Purdue University
West Lafayette, IN

Identifying the need for a multi-faceted, interdisciplinary research team to investigate various issues within the aviation industry, Purdue University's Department of Aviation Technology has organized the Advanced Aviation Analytics Institute for Research—A Center of Research Excellence (A^3IR-CORE). A thorough meta-analytic review of collaborative network theory was the impetus for the organizational development of this forum. Focused on engaging a wide range of knowledge, talents, and experiences, this preeminent institute will fill the traditional void of academic material in the aviation industry by actively involving undergraduate students, graduate students, and faculty in a collaborative environment that increases the dissemination of research materials within the college and expands departmental recognition within the larger discipline. The flagship function of this group is to facilitate and manage research projects for departmental faculty—projects that will lead to publication at nationally and internationally recognized conferences and in peer-reviewed journals.

Scholarly progress hinges on the ability of researchers to establish connections and coordinate activity. The aviation community has in particular faced challenges in this area due to the large number of both academics and practitioners spread over various areas of the country. The Advanced Aviation Analytics Institute for Research–a Center of Research Excellence (A^3IR-CORE) was established to foster a collaborative environment between scholars at all levels. The support of faculty research projects and publications are the flagship offerings of this group, leading to recognition in academic journals and at leading conferences around the world. Alluding to various existing scholarly publications, this paper intends to build support for the inception of a collaborative research group within the Department of Aviation Technology at Purdue University

Collaborative Network Theory

Ad-hoc collaborative networks of isolated researchers with other isolated researchers are the backbone of much scholarly research, but opportunities for advances can be lost without a formalized structure for cooperative work and information sharing. The creation of more formalized networks for research dissemination has proved beneficial in the past, notably with Saint Louis University's Safety Across-High Consequence Industries (SAHI) conference (Bowen, Block, & Patankar, 2009). The concept emerged as a group of interested individuals sharing information; the creation of a conference to facilitate this sharing process allowed many others to also benefit from the initial gains achieved in safety through such dissemination.

Formalized structures ensure that research ideas and techniques are captured and available for others to use. As described in Bowen and Lu (2004), this process enables synergistic relationships between researchers, regardless of their proximity.

Void of Academic Material

Aviation has traditionally struggled to find an appropriate outlet for collaboration on high-level research projects (Bowen, et al, 2011). A^3IR-CORE aims to provide an avenue for faculty and students (both undergraduate and graduate) to submit academic work for widespread dissemination. The addition of a duo of open-access, scholarly resources to the existing research lab concept will increase access to the generated scholarship.

The first of these resources is Purdue e-Pubs, a subsidiary of Purdue University Press (PUP). E-Pubs is an open-access repository that allows scholars to archive their work in a database that is accessible through Google Scholar. A^3IR-CORE administers the various Aviation Technology series; although submissions are not peer reviewed, it should be noted that publications are screened prior to submission, to ensure the works are of appropriate academic quality and reflect well on the department and aviation industry.

Secondly, the Journal of Aviation Technology and Engineering (JATE) is also operated and published by A^3IR-CORE staff members. This biannual journal draws global submissions focused on issues that surround the aviation and engineering industries; a major function of JATE is to "promote the bridging" of these two areas in a collaborative environment. Through a double-blind, peer-reviewed process, the journal provides an academically rigorous forum for researchers, scholars, and practitioners to distribute their work to other researchers.

Collaborative Environments

One of the aspects most critical to the lab's survival is the cohesive involvement of students from various scholastic levels combined with vital faculty mentoring. From the early stages of the lab's development, it has been imperative to recruit diverse student participation among continuing scholars including doctoral candidates and master's students. Then throughout its expansion, the lab began outreach to effervescent undergraduate researchers as an investment in future and long term goals while promoting continued schooling in higher education. These combined levels of education foster a conglomeration of ideas and experiences, which motivates and inspires cutting-edge research and development. This progression is then supervised by faculty advisors with distinct knowledge and experience within various facets of the aviation industry. Together they form a professional research group capable of tackling various problems and issues from industry.

Team members and faculty advisors bring new research topics and suggestions to weekly meetings, which facilitates collaboration and constructs innovative projects. Ultimately, it is this participation that builds the foundation for the future of the lab by producing published research papers and articles. Multiple studies have shown that this diverse enterprise among students and faculty encourages progress and achievement. Dotterer (2002) asserts, "Any campus that motivates its students to learn through individual and collaborative research—and can find ways to support these intellectual journeys with the necessary human and material resources—certainly does provide its students with a first-rate education" (p. 82). This ideology is crucially implemented in the foundation of A3IR-CORE and supports current and future progress among all members and advisors.

Under the supportive mentoring and guidance of faculty advisors, team members flourish in research opportunities and continuously cultivate advanced strategies. Additionally, a multitude of available technical resources also facilitate research within the lab. The lab consists of several computers with statistical analysis capabilities as well as various reference books to ensure accurate and efficient production. Faculty advisors uphold further resources in the industry for additional assistance and reference. Furthermore, the combination of the lab's resources, references and knowledge cultivates a dynamic research environment inspiring active participation and original conceptions.

Organizational Structure

Since the lab's inception in May 2012, a vertical organizational structure has emerged as the preferred hierarchy; as time progressed, additional levels and layers of staff have been added to create a dynamic group with diverse skillsets and expertise. The Team Members make up the core of the research group, assisted by Team Volunteers. Both of these groups report to the Team Leader. Directly above the Team Leader is a horizontal level consisting of the Managing Director, Global Research Scholars, and Faculty Advisory Board. The aforementioned groups have direct access to the Director for collaboration and research support. A visual depiction of the organizational structure appears on the next page (Figure 1). Solid lines depict the research lab's chain of authority; broken lines, on the other hand, are indicative of a collaboration relationship that has access to various parties for research ideas, support, etc.

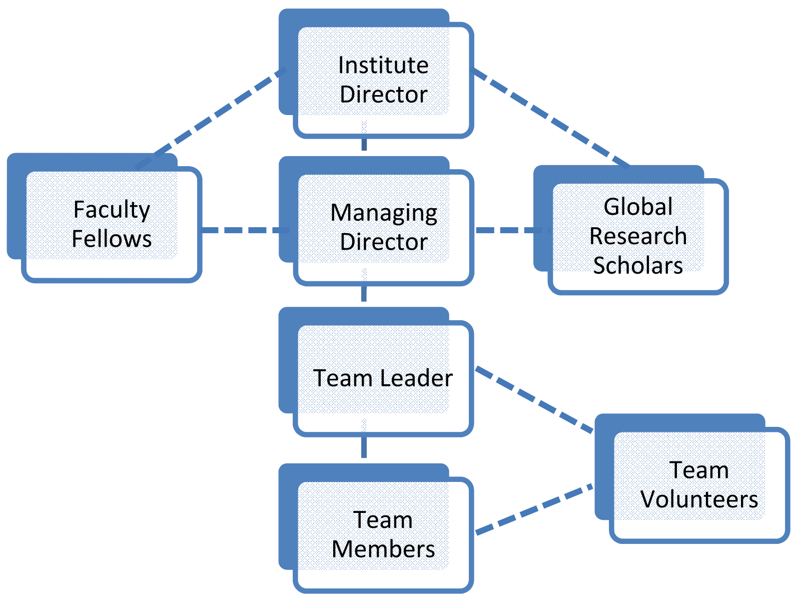


Figure 1. A[3]IR-CORE Organizational Chart

Director—the head of the research group, the director sets the pace and priorities of the lab. Allocation of funds, approval of research activities, and final decision-making and delegation authority rests at this level. Collaborating and building relationships with external faculty members and scholars is an important role that ensures a well-rounded membership.

Faculty Fellows—internal faculty members make up the core research support function of the group; providing research ideas and expertise in their respective areas, this board acts as a catalyst to start research in new areas otherwise left unexplored. Active participation from these members is cyclical, and often corresponds to the projects at hand.

Global Research Scholars—working collaboratively with academics and industry leaders from other universities and various entities abroad, global research scholars enrich and diversify the experience and expertise of the research lab. Building bridges across the country and around the world is paramount in creating and maintaining relationships that foster greater learning potential, and increased exposure to emerging research ideas.

Managing Director—working closely with the team leader, faculty advisory board, global research fellows, and director, the managing director retains responsibility for all lab functions. This positions acts as a mentor to all lab members, while spearheading projects and giving final approval before publications move from the research lab to the next step in the publication process. Maintaining a healthy and active relationship with internal and external contacts, the managing director represents the lab at various functions, both domestically and abroad.

Team Leader—the team leader serves to increase cohesiveness between group members, manage the lab's continuous operations, and act as an efficient liaison between the research group and upper management (managing director and director). Scheduling staff, preparing for and leading working sessions/meetings, and acting as the general research group "organizer", the team leader has a continually evolving role within the group, and must be ready to adapt to the dynamic environment

Team Members—after serving one academic semester as a team volunteer, interested parties are eligible for promotion to the team member position; compensation for time may include course credit, funding, or hourly pay. Promotion from volunteer status includes an increase in responsibility, including time commitments, publication output, and general responsibility for completion of projects, both individual and in tandem with the larger research group.

Team Volunteers—team volunteers are entry-level researchers looking for an area of interest to explore; appointment occurs after an informal submission of a resume/CV followed by an internal review during a regular ream working session. A standardized letter of agreement is endorsed by the applicant and managing director upon recommendation by the team leader and concurrence of the team members.

Table 1.
Current A3IR-CORE Lab Staff Roster

Director		
Brent D. Bowen, Ed.D.	Professor and Department Head, AT*	Purdue

Faculty Fellows		
Erin E. Bowen, Ph.D.	Assistant Professor, TLI**	Purdue
Thomas Carney, Ph.D.	Professor, AT	Purdue
Chien-tsung Lu, Ph.D.	Associate Professor, AT	Purdue
John Mott, M.S.	Assistant Department Head, AT	Purdue
Michael Suckow, MBA	Assistant Department Head, AT	Purdue
Stewart Schreckengast, Ph.D.	Associate Professor, AT	Purdue

Global Research Scholars		
Mary Fink, M.S.	Academic Consultant, AT	Purdue
Nanette Metz, Ph.D.	Adjunct Faculty	Embry-Riddle
Allen Hamilton, Ph.D, G.G.	Lecturer, AT	Purdue

Core Research Group		
Jennifer Kirschner, M.S.	Managing Director, A^3IR-CORE Research Lab	Purdue
Clay Wildt, B.S.	Team Leader, A^3IR-CORE Research Lab	Purdue

Reilly Meehan, Student	Team Member, A3IR-CORE Research Lab	Purdue
Branden Avery, B.S.	Team Member, A3IR-CORE Research Lab	Purdue
Anthony Erstad, B.S.	Team Volunteer, A3IR-CORE Research Lab	Purdue

Note. *Aviation Technology, **Technology Leadership & Innovation

Current Research

Utilizing the multitude of expertise and resources available to our newly conceptualized research lab, various departmental projects will be routed through the lab for management and eventual completion. Using the collective knowledge of qualitative and quantitative methodology developed through intense and focused training, members of the research group will be prepared to tackle projects in the areas of public policy, airline quality measurement, passenger perceptions of airline quality, economic indicators of airline performance, management style, as well as a host of other airline-related topics.

Additionally, the lab currently partners with Purdue University Libraries to host an open-access repository accessible to faculty and staff within the Department of Aviation Technology; through this service, scholars at all academic levels will be able to archive their work, regardless of the format. For instance, a student would be able to submit a poster presented at conference proceedings, PowerPoint slides used during a presentation, or white papers before they are submitted to other venues. By using analytic software built into the repository system (e-Pubs), authors can track the download count and geographic location of interested parties. Advocating for and administering the e-Pubs process within our own department, we hope to increase scholarship and publication by facilitating a seamless avenue to sharing work via the repository.

Departing from the traditional quantitative-centric view of research, the research group has worked hard to properly train and equip members of the lab to embrace other methodological techniques including qualitative data analysis. Using software such as NVivo 10, scholars have the opportunity to approach research from alternative angles using nontraditional data such as interviews, photographs, videos, essay-type survey responses, etc., to fully quantify previously unquantifiable data. In the aviation industry, the ability to adapt and use new techniques to investigate problems will produce innovative solutions.

Future Impacts

The ever increasing pace of globalization means that our economy is inextricably linked to the world. As more countries increased their rate of development and gain access to the middle class, the demand for air travel has skyrocketed in Asia, Africa, and Latin America. The growing pains associated with any rapid industrial expansion necessitates a strong focus on quality assurance and standards in order to ensure that the highest levels of safety are maintained. Airlines, and more broadly, aviation policymakers, require access to academic sources of information about best practices; this new demand for information access will need to be facilitated properly in order to ensure that information distribution is as streamlined as possible. This opens the door for researchers with expertise and global reach to link up with other researchers and industry partners around the world.

Acknowledgements

The authors would like to express their gratitude to all those who gave us the possibility to complete this report. We also want to acknowledge the invaluable assistance of the Applied Human Factors Research Laboratory and the Advanced Aviation Analytics Institute for Research—A Center of Research Excellence (A3IR-CORE); without these vital research groups, this paper would not have been possible.

References

Bowen, B. D., Block*, E. E., & Patankar, M. (2009). A network collaborative design construct for the dissemination of aviation safety research. In *Proceedings of the 15th International Symposium on Aviation Psychology.* Dayton, OH: Association for Aviation Psychology.

Bowen, B. D., Bowen, E. E., Lehrer, H. R., Mott, J. H., Watkinson, C. T., Newton, M. P., & Kirschner, J. E. (2011). The digital migration of research dissemination in aviation psychology disciplines. In *Proceedings of the 16th International Symposium on Aviation Psychology.* Dayton, OH: Association for Aviation Psychology.

Bowen B., Patankar, M., Block, E. (December 2008). *White Paper on Development of the National Center for Aviation Safety.* Saint Louis University. 22pps.

Bowen, B., & Lu, C-t. (2004). Proposing a comprehensive policymaking mechanism: The introduction of policy research construct (PRC). International Journal of Applied Aviation Studies, 4(1), 31-44.

Dotterer, R. L. (2002). Student-faculty collaborations, undergraduate research, and collaboration as an administrative model. *New Directions for Teaching and Learning, 90*, 81-8

---

*E. E. Block is now E. E. Bowen and has published scholarly articles in the aviation and psychology literature under both names.

SYNTHETIC TASK ENVIRONMENTS AND THE THREE BODY PROBLEM

John M. Flach
Wright State University
Dayton, OH 45431 USA

The challenge for our panel was to address the opportunities and challenges of synthetic task environments for basic research on human performance in sociotechnical systems. In doing this, the classical three-body problem from physics is used as a metaphor to illustrate the contrast between dyadic and triadic semiotic models of cognitive systems. In the context of this metaphor, synthetic task environments offer a means to bring some of the additional complexities of triadic semiotic systems under experimental control where converging empirical methods can help to titrate through the additional complexity to distill basic theoretical insights that will potentially have practical value for training and interface design.

This paper will begin by examining two alternative perspectives on semiotic or cognitive systems – the dyadic and the triadic perspectives. The case will be made that the triadic perspective provides a more comprehensive framework for exploring cognition and the pragmatic implications for the design of sociotechnical systems (e.g., next generation air space control systems). However, the introduction of a 'third body' into the cognitive system raises important challenges for both science and application. The paper will consider the challenges of the triadic (three-body) system and will suggest how synthetic task environments can help researchers to address these challenges.

Semiotics

The theoretical context for cognitive science and for its application to the design of sociotechnical systems was strongly influenced by the field of semiotics. Semiotics is typical described as the science of signs, but it can also be described as the science of meaning making. That is, the focal question of semiotics is how meaning is attributed to signs or representations. Ferdinand Saussure and Charles Sanders Peirce are typically credited with independently founding the field of semiotics (Eco, 1979, Morris, 1971). However, they approached the problem from two distinct perspectives.

Sausure's Dyadic Semiotic System

Saussure, generally regarded as the father of linguistics, framed the semiotic system in terms of the dyadic relation between a sign/symbol and an agent/observer, as illustrated in Figure 1. Saussure's interest was particularly in the evolution of alphabets and languages. Thus, he viewed the semiotic problem from the perspective of assigning meaning to symbols (e.g., written or spoken language). This framework fit ideally with the computer metaphor of mind and it set the stage for the first wave of cognitive science and the information processing approach to cognition and design. In this context, the cognitive agent was considered to be a symbol processor and the focus of basic research was on exploring the internal information processing constraints (e.g., channel capacity and internal recoding). The focus for application of this approach involved characterizing the internal information constraints so that these constraints could be considered in designing cognitive work (e.g., don't overload the limited capacity working memory).

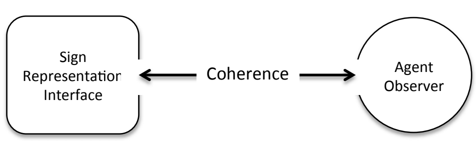


Figure 1. This diagram illustrates Saussure's dyadic model of the semiotic system.

In applying the dyadic approach to sociotechnical systems, it was natural to focus on the coherence between the surface structure of the interface (i.e., the symbol or representation) and the responses or interpretations of the human operator. Research hypotheses in this paradigm were typically framed in terms of the coherence between general surface properties of the interface and the information processing demands. Classical examples include early work on shape coding to improve discriminability among different controls (Jenkins, 1947) and work on stimulus-response compatibility that looked at the coherence of the spatial topology of the display representation relative to the spatial topology of the response (Fitts & Seeger, 1953). More recently, attention has been given to the organization or clustering of information in the display (e.g., intergral versus separable displays), relative to hypothetical information processing limitations (e.g., parallel versus serial processing) (Wickens & Carswell, 1995). In all these instances, hypotheses about the relative effectiveness of alternative representations were often tested using generic tasks motivated by assumptions about the relevant information processes.

Peirce's Triadic Semiotic System

Peirce, the father of Pragmatism, was interested in the pragmatics of belief and action in the world. How is it that our beliefs about the world can become the basis for successful action in the world? Thus, Peirce brought a third component into the semiotic system. In essence, the third component reflects a source behind the sign or representation – i.e., a problem domain or a natural ecology. By adding this third component, Peirce brought two additional relations into the semiotic system. In addition, to the *coherence* between the sign and the expectations of the agent considered in the dyadic system, the triadic system involves the *structural mapping* between the sign and the source domain and the *correspondence* between the agent's beliefs about action and the actual consequences of action in that source domain as illustrated in Figure 2.

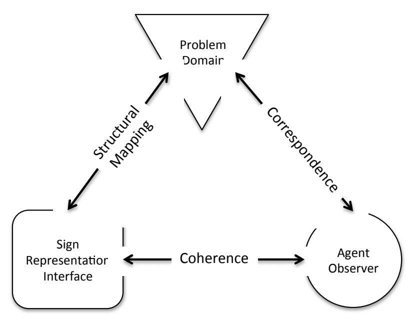


Figure 2. Peirce's triadic model of semiotics introduces a third 'body' into the system.

In the triadic model, the semiotic problem changes from interpreting a symbol to adapting to the demands of a problem domain. Rather than the symbol or representation being the 'stimulus,' it becomes simply a medium, with the stimulus displaced to the problem domain. The ultimate test of the triadic system is not whether the representations match the agent's expectations and beliefs, but rather whether the agent's expectations and beliefs support successful interactions with the problem domain. Attention shifts from the *syntax* of the surface features of the interface representation to the *semantics* associated with the deep structure of the problem domain. And the pragmatic design goal is to shape the agent's expectations through training and/or interface design in ways that lead to productive interactions with the problem domain.

Note that in the triadic semiotic system, the *USER-centered* concerns associated with the coherence between the interface and agent expectations remain an important component of the semiotic system. However, the triadic model also raises additional *USE-centered* concerns about the relations between structure in the representation and the functional constraints associated with the target problem domain (Flach & Dominguez, 1995). *In the context of the triadic model, the design challenge shifts from* *'matching' the agent's mental model, to 'shaping' the agent's mental model so that it supports productive action with regards to a target problem domain*.

The Three Body Problem

As physicists know, modeling the motion of interacting bodies in space becomes significantly less tractable when a third body is introduced. This is one of the major attractions of the dyadic approach to semiotics. Using the dyadic framework the image guiding research was that of a communication channel and problems of cognition were reduced to open-loop, symbol processing problems, constrained only by internal information processing limits as illustrated in Figure 3A. In this context, research questions became significantly more tractable in terms of identifying simple causal relations between stimuli and responses. This allowed the use of simple laboratory paradigms motivated by information processing models for independent stages of processing. The general stage specific tasks required no special knowledge so that general populations of readily accessible participants could be studied. Thus, large-N studies were feasible and it was possible to use strong statistical inference to judge effects.

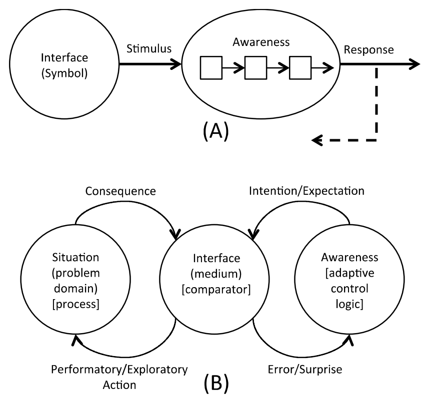


Figure 3. This diagram illustrates that introducing a third body changes the underlying dynamics from open-loop (i.e., causal) to a closed-loop (i.e., self-organizing).

In contrast to the communication channel metaphor, the triadic model of semiotics suggests a dynamic closed loop coupling between perception and action as illustrated in Figure 3B. This reflects an abductive logic where the 'tests' of beliefs are the practical consequences from acting on those beliefs. In this dynamic, the 'sign' interface has a dual function in terms of action/control (i.e., comparing the difference between consequences and intentions – *error*) and perception/observation (i.e., comparing the difference between consequences and expectations – *surprise*). This leads to a self-organizing dynamic where the cognitive agent is simultaneously shaping actions and being shaped by the ecological consequences of those actions. To understand the dynamics of the triadic system it becomes necessary to understand the constraints associated with the work domain or problems space (i.e., deep structure) and the potential interactions of these constraints with the internal constrains (i.e., mental models) of the human agents in relation to observation and control. In the following sub-sections some implications for approaching the triadic semiotic system are considered.

Cognitive Task versus Work Domain Analysis

As reflected in the images of the triadic semiotic system, a necessary step in a triadic approach is to bring the work ecology into the research frame. Thus, a prerequisite is to identify the deep structure of that ecology. This is the goal of Work Domain Analysis (WDA) (Vicente, 1999). To set the context for this, it is important to distinguish WDA from Cognitive Task Analysis (CTA) (e.g., Fleishman & Quaintance, 1984). CTA has typically been designed to reflect the information processing activities associated with the work. This makes perfect sense from the dyadic perspective where the focus was on cognitive activities inside the head of the human agent. In contrast, the focus of WDA is on the functional constraints associated with the problem domain. For example, in aviation this includes the aerodynamic constraints on vehicle motion, situational factors within the airspace (e.g., weather), the regulatory constraints on airspaces, as well as the value constraints against which safety and efficiency are measured. The goal is to better understand the 'deep structure' of the problem.

Representative Design of Experiments

Research motivated by the dyadic approach is often designed to isolate variables associated with specific internal information processing stages. Thus, the choices of tasks and independent variables are typically motivated by models of the internal stages. Even when the research is conducted within high fidelity simulations (e.g., a flight simulator), the research will often focus on secondary tasks (e.g., memory search or probe reaction time) that are thought to tap into the relevant internal mechanisms.

In a triadic approach, however, the focus is on how performance is shaped by the deep structure of the problem domain. Thus, the tasks and independent variables are explicitly chosen to reflect that deep structure. This requires that the evaluation context be *representative* of the work domain. It is important to appreciate that representativeness does not simply mean that the interface (e.g., knobs and dials) functions properly (e.g., as in a high fidelity flight simulation). It also involves the validity of the problem that is driving the interface – that is the dynamics of the problem context that the research is intended to generalize to. So, for example, the triadic approach requires the experimental *situations* or context are representative of the target domain. For example, in evaluating a design of new technologies for the next generation of air space management systems, it would be important that the evaluation contexts involve conditions that would be representative of future flight conditions (e.g., in terms of air traffic densities and regulatory constraints).

In addition to care in selecting the experimental task scenarios, it also becomes important to select participants from representative populations. For example, one cannot simply select a participant from an Introductory Psychology course and expect him to be able to fly a simulated aircraft under realistic air traffic conditions. Thus, the triadic approach demands care in selecting participants who have the appropriate skills and experience to address the problems presented. This raises the issue of competency.

Mission Essential Competencies

The construct of *Mission Essential Competencies* has emerged in the context of training applications and research (Alliger, Beard, Bennett, Colegrove, & Garrity, 2007). In contasting the dyadic and triadic approaches, the key distinction reflected in this construct is a shift from focusing on generic information constraints to focusing on "mission relevant" abilities, skills, experience, and knowledge. Thus, the construct of competencies focuses on the deep structure of work in terms of demands for success in a specific work domain. For example, with respect to air combat, Colegrove and Alliger (2002) define MEC as "higher-order individual, team, and inter-team competency that a fully prepared pilot, crew, flight operator, or team requires for successful mission completion under adverse conditions and in a non-permissive environment." In essence, constistent with the triadic approach, the MEC construct *situates* or grounds the properties of the cognitive agent (i.e., awareness) relative to specific demands of a work domain (i.e., situations) and this provides a triadic basis for making decisions for designing training scenarios and goals.

Ecological Interfaces

Training reflects one path for shaping the internal models of operators so that they better correspond with the deep structure of specific problem domains leading to more productive actions. Another means for shaping the internal models of operators is through the design of interface representations (Bennett & Flach, 2011; Rasmussen & Vicente, 1989). The construct of *Ecological*

Interface Design (EID) provides a triadic alternative to the conventional dyadic approach that tends to emphasize matching generic internal models (e.g., population stereotypes), rather than shaping internal models so that they better correspond with the demands of specific work domains. The emphasis of the EID approach is on designing display constraints (e.g., configural visual graphics) that are explicitly mapped to the underlying deep structures of the work domain. In this context, the emphasis shifts from focus on capacity limitations to skills such as chunking that allow experts to by-pass these limitations in order to meet the demands of complex tasks (e.g., Chase & Simon, 1973; Ericsson & Charness, 1994). For example, research on chess suggests that the ability of chess experts to remember board positions and to quickly focus on good alternative moves reflects a different way of chunking information. Novices focus on individual 'pieces' and experts focus on the spaces that the pieces are attacking (Reynolds, 1982). Thus, structure in configural graphics is designed to bias operators toward organizing (i.e., chunking) information in ways that support productive thinking or expertise.

Synthetic Task Environments

The previous section illustrated some of the ways that the addition of the third 'body' to the semiotic system changes the questions that become most interesting for researchers. The clear implication of this shift for research is that it becoms necessary to incorporate the deep structure of specific work domains into the experimental contexts. Fortunately, information technolgies such as high fidelity simulators and virtual environments provide one means to do this. These technologies allow researchers to build *synthetic enviroments* that represent the deep structures of specific work domains with a relatively high level of fidelity. While bringing more of the richness of natural work domains into the laboratory these synthetic environments offer possibilities for manipulation and replication of conditions that would not be possible in naturalistic settings. Additionally, these environments typically allow unobtrusive measurement of both the situation (i.e., independent variables) and operator performance (i.e., dependent variables) in ways that often are not possible in natural settings.

The Measurement Problem

The ability to simultaneously measure properties of the changing situation and the performance of operators at multiple levels of abstraction is both the biggest opportunity and the biggest challenge of synthetic task environments. On the opportunity side, one of the biggest challenges for conventional research focused on generic information processing tasks was to relate statistically significant differences observed in laboratory tasks to practical differences in specific work domains. Would a significant laboratory effect on reaction time translate to a practical difference in operational effectiveness? Synthetic task environments provide a means to address this question empirically. That is, within a synthetic task enviroment it is possible to simultaneously measure micro-level performance differences (e.g., reaction time to a specific display event) and more macro-level functional differences (e.g., winning or losing an engagement).

Comparisons across levels of abstraction provide empirical evidence about whether differences at the micro-level are correlated with success at the macro-level. Thus, questions about operational implications can be answered based on empirical evidence at the operational level and patterns between this evidence and other variables that might be more closely related to generic and specific constraints associated with internal mental models. Such measurement opportunities can provide a bridge between practice and theory that will lead to improvement on both ends. This bridge is particularly important for complex, nonlinear systems where analytical linear extrapolations fail, and insight typically depends on empirically linking quantitative changes at the micro-level with qualitative changes at the macro-level (e.g., Shaw, 1984).

The biggest challenge for research using synthetic task environments is data overload. The opportunity to measure everything, can make it harder to see anything. Based on my own experiences, I venture the guess that many research programs using synthetic task enviroments have oodles of data that get archived, but that are never analyzed or examined. In order to take advantage of the data that synthetic environments make available to researchers, it can be essential that the search of that data is guided by theories about the deep structure of the work domain, about the domain specific competencies required, and about the generic constraints on awareness. The three body problem is inherently intractable! Thus, solution depends on clever partitioning of the problem and the use of converging operations to discover and isolate signals (e.g., patterns associated with fundamental properties) that are embedded in the complexity.

References

Alliger, G.M., Beard, R., Bennett, W., Colegrove, C.M. & Garrity, M. (2007). *Understanding Mission Essential Competencies as a workload requirement.* Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division. AFRL-HE-AZ-TR-2007-0034.

Bennett, K.B. and Flach, J.M. (2011). *Display and interface design: Subtle science, exact art.* Boca Raton, FL: CRC Press.

Chase, W.G. & Simon, H.A. (1973). The mind's eye in chess. In W.G. Chase (Ed.). *Visual information processing.* New York: Academic Press.

Colegrove, C.M. & Alliger, G.M. (2002). Mission Essential Competencies: Defining combat mission readiness in a noval way. Paper presented at the NARO RTO Studies, Analysis and Simulation Panel (SAS) Symposium. Brussels, Belgium. (April).

Eco, U. (1979). *A theory of semiotics.* Bloomington, IN: Indiana University Press.

Ericsson, K.A. & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist,* 48, 725-747.

Fitts, P.M. & Seeger, C.M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology,* 46, 199-210.

Flach, J.M. & Dominguez, C.O. (1995). Use-centered design. *Ergonomics in Design*, July, 19 - 24.

Fleishman, E.A. & Quaintance, M.K. (1984). *Taxonomies of human performance: The description of human tasks.* Orlando, FL: Academic Press.

Jenkins, W.O. (1947). The tactual discrimination of shapes for coding aircraft-type controls. In P.M. Fitts (Ed.) *Psychological research in equipment design.* Army Air Force, Aviation Psychology Program, Research Report 19.

Morris, C. (1971). *General theory of signs.* Paris: Mouton.

Rasmussen, J. and Vicente, K. (1989). Coping with human errors through system design: Implications for ecological interface design. *International Journal of Man-Machine Studies*, 31, 517-534.

Reynolds, R.I. (1982). Search heuristics of chess players of different calibers. *American Journal of Psychology,* 95, p. 373-392.

Shaw, R. (1984). *The dripping faucet as a model chaotic system.* Santa Cruz, CA: Arial Press.

Vicente, K.J. (1999). *Cognitive work analysis: Toward safe, productive and healthy computer-based work.* Mahwah, NJ: Erlbaum.

Wickens, C.D. & Carswell, C.M. (1995). The proximity compatibility principle: Ist psychological foundation and relevance to display design. *Human Factors,* 37(3), 473-494.

EXTENDING MISSION OPERATIONS SAFTY AUDITS (MOSA) RESEARCH TO AN INDIAN SUB CONTINENT ISLAND AIRLINE

Sue Burdekin
University of New South Wales
Australian Defence Force Academy
Canberra, Australia

The aim of the Mission Operations Safety Audit (MOSA) research is to validate behavioural self-reported data from professional pilots, so that management can have confidence in this safety-critical debriefing information, and feed it back into the training continuum. In doing so, a safety loop can be established in a cost effective, operationally specific and timely program of data collection. The first study was conducted in a military F/A-18 Hornet simulator. Pilots were asked to self-report on their own operational performance across a predetermined selection of behavioural categories designed in conjunction with subject matter experts. To further test the MOSA methodology, this time in-flight, a second study was carried out with the cooperation of a civil airline in Europe. Both the military and the civil airline studies found that professional pilots were able to effectively self-report on their own performance. However, the multi-crewed European airline pilots' results revealed that first officers were more critical of their own performance. In order to determine whether national or organisational culture influenced these results, the MOSA methodology was recently tested in a regional airline operating Dash 8 aircraft between island destinations in the Indian Ocean. The results indicated that neither national culture nor aircraft operating culture appear to influence the accuracy of pilot self reports. However, the self reports from first officers do appear to be linked to their seniority and experience in that role.

The use of trained observers to rate professional pilots' non-technical skills performance during normal flight operations has become an accepted individual assessment and system evaluation method in civil aviation. There are two well know approaches utilizing in-flight observers: NOTECHS (non-technical skills) assessment and Line Operations Safety Audits (LOSA). NOTECHS was developed from a Joint Aviation Authorities (JAA) sponsored research project in the UK and Europe to provide a tool to assess pilots' CRM skills in the cockpit. These skills were defined as cognitive and social skills not directly concerned with flight control, systems management and standard operating procedures (SOPs). As an assessment tool supported by the regulator, failure to meet the predetermined standard can result in license suspension and remedial training (Flin, O'Connor & Crichton, 2008). On the other hand, LOSA was developed by The University of Texas Research Project and airline partners in North America initially to audit pilots CRM performance and then expanded to identify threats to the conduct of the flight and how errors are managed. It provides a report of an airlines strengths and weaknesses determined by non jeopardy observations of line crew performance on the flight deck (Klinect, Murray, Merritt & Helmreich, 2003). In-flight observations for both NOTECHS and LOSA are conducted under 'normal' conditions. That is, during regular, revenue raising operations.

However, the definition of ‘normal operations’ in a military context translates into a completely different paradigm. Normal military operations can involve high speed, rapid manoeuvre, or terrain following radar flight in a range of multi-crewed and single pilot aircraft. Many of these aircraft types can only accommodate the flight crew therefore, depending upon the particular platform, in-flight observation may not be physically possible. Because of this, the Royal Australian Air Force (RAAF) supported research into testing the accuracy of pilot behavioural self-reports, known as Mission Operations Safety Audits (MOSA). The objective was to establish a valid means of collecting targeted safety critical debriefing information which could be acted upon in a timely manner. Additionally, it was intended that MOSA could provide a cost effective evaluation of aircrew performance which would highlight issues that could feedback into the training continuum, thereby establishing a behavioural safety loop in line with the defence aviation safety management system.

The first study was conducted using a single pilot F/A-18 Hornet flight simulator. Participants were asked to rate their own behaviour on a 5 point Likert scale after ‘flying’ a medium or high workload mission, according to pre-determined behavioural categories, which had been designed with input from military subject matter experts (SMEs). These categories included: workload management, communication, evaluation of plans, situational awareness, monitoring and cross checking, inquiry, assertiveness and automation management. The results across both conditions of workload were highly correlated with the ratings of expert observers (Burdekin, 2012).

Following the military simulator research, Airbus supported a second MOSA study to test self-report methodology on the flight deck during normal operations in a European multi-crewed civil airline environment. SMEs from Airbus and the ‘low cost model’ airline Easyjet joined the researcher to further develop the methodology to incorporate a crossed design which recorded self reports from each pilot, along with the captains ratings for the first officer (F/O) and the F/Os ratings for the captain. These ratings were then compared to the ratings from an observer during the same flight sector. Sixty flight sectors were observed and the ratings by the observers were found to be significantly correlated with the self-report ratings from both the captain and the F/O. However, deeper analysis revealed that F/Os were more highly critical of, not only their own performance, but also the performance of their captain (Burdekin, 2012). It was noted that 83% of the F/Os who volunteered for the study were relatively new recruits of less than one to three years with the company and an average of 1950 hours total flight time. Given the current debate concerning the minimum flight hours required by pilots to fly for an airline (US Congress, 2010; Rural Affairs and Transport References Committee, 2011), the MOSA research has highlighted that junior F/Os expect higher standards of performance from themselves and their senior crew members. In order to determine if the level of experience and/or perhaps a cultural issue influenced the European civil airline study results a further test of the MOSA methodology was conducted.

The present study was carried out in an Indian sub continent airline operating Dash 8 turbo-prop aircraft to island holiday destinations in the Maldives. The aim of the study was to determine if the MOSA methodology would be suitable for collecting accurate aircrew self-reported information in a different cultural and operational environment.

Method

Forty one flight sectors were observed from the jump seat during normal revenue raising operations by an experienced subject matter expert (SME) who is also a commercial pilot. All crew members were volunteers and their participation was anonymous. As such, it was not possible to determine if a pilot was observed more than once. The design of the study was developed by the researcher with input from the safety manager, senior management and check and training pilots. A crossed experimental design was developed where each volunteer crew member rated his/her own performance, the performance of the other crew member and how, collectively, they performed as a crew during each sector, across a predetermined set of behavioural categories. The observer rated the performance of the captain, the first officer and also how they performed as a crew during each sector using the same predetermined behavioural categories. The protocols included eight categories of behaviour that were assessed to be a representation of the non-technical skills that the airline was keen to evaluate. Those categories were: briefing; contingency management; monitor/cross-check; workload management; situational awareness; automation management; communication; and problem solving/decision making. Each behavioural category was given comprehensive descriptors, illustrated in figure 1.

Figure 1.
Example of behavioural category and descriptors

BEHAVIOURAL CATEGORY	**DESCRIPTOR**	**GRADING/WORD PICTURE** *(1. Poor; 2. Marginal; 3. Adequate; 4. Very Good; 5. Excellent)*
Automation management	**Interaction between the operator and automated system**	**1. Incorrect crew interaction and management of aircraft automatic systems. Clear errors of competency in automation set-up, mode selection and utilization.** **2. Basic interaction with aircraft automatic systems. Appropriate mode selection and utilization barely adequate to maintain safe flight profiles.** **3. Level of automation interaction adequate to maintain prescribed SOP profiles. Mode utilization satisfactory and procedurally correct. Recovery technique from anomalies reflects limited system awareness.** **4. Automation interaction to a good standard. Effective and timely management of automatic modes. Flight path SOP profiles maintained to a proficient standard. Clear understanding of aircraft automation systems reflected in sound anomaly management.** **5. Automation management to a high standard. Clear anticipation and use of appropriate modes. All anomalies managed to a highly proficient standard reflecting a deep understanding of the automation system.**

The protocol form allowed for crew members and the observer to comment on any issue that affected the safety of that sector in the form of air traffic management, ground support, aerodrome operations, cabin crew interaction, and any other issues.

Additionally, a questionnaire was conducted during the flight that asked crews to identify any wider safety issues of concern within the airline. For example: "Can you list the top 5 safety issues currently affecting the company"; "Can you predict what the next incident/accident will be"; "What do you think would be the best way to prevent this from occurring"; "Can you nominate one CRM strategy that you have adopted that has changed the way you approach your flying"? The volunteer crew members were assured of individual confidentiality and the researcher/observer remained the 'gate keeper' of all data.

Results

The data across all categories of behaviour were collapsed and subjected to a test of correlation between the ratings of the independent observer, the captain, and the first officer (refer to Table 1). In addition to rating the performance of themselves and each other, the crew members were asked to rate how they performed together as a crew. These ratings were also correlated with the ratings of the crew's performance from the independent observer (refer to Table 2). All results were found to be statistically significant except for the crew performance ratings from the first officers. Details of crew experience can be found in Table 3.

The answers from the in-flight questionnaire were compared with routine safety data gathered from flight data monitoring, voluntary and mandatory safety reports, air safety occurrences, management safety committee meetings, and other workforce/organisational evaluation data. Analysis of this sensitive safety critical information indicated that the results were valid.

Table 1.
Results across all categories of behaviour

Rater	Mean	sd	N	r
OBS	4.12	.54	224	.669 **
CAPT	4.15	.62		
OBS	4.13	.53	224	.188 **
F/O	4.13	.58		
CAPT	4.14	.68	224	.232 **
F/O	4.13	.58		
F/O	4.31	.62	224	.112 *
CAPT	4.15	.62		

** Significant .01
* Significant .05

Table 2.
Crew performance results

Rater	Mean	sd	N	r
OBS/CREW	4.25	.52	224	.620 **
CAPT/CREW	4.26	.58		
OBS/CREW	4.25	.52	224	.186 **
FO/CREW	4.29	.59		
CAPT/CREW	4.26	.58	224	.099
FO/CREW	4.29	.59		

** Significant .01
* Significant .05

Table 3.
Average age and experience of Crew

	Mean Age	Average Total Flight Hours
Captain	46	14,200
First Officer	31	4,300

Conclusion

The results from this study show that both captains and first officers were able to accurately report on their own performance, compared to the ratings from each other and an independent observer, across a range of categories of behaviour which reflected their non-technical skills. Although the observer/captain ratings were more highly correlated in both this and the European studies, in the present study, the anomaly of first officers being more critical of their own performance than the ratings issued to them by their captains and the observer was not repeated. The difference in the self-assessment of these first officers might be explained by their level of experience and the length of time that they had been employed as first officers. The majority of the F/O volunteers in this study were senior first officers who were awaiting a captaincy slot. Whereas, the first officers in the European study were relatively junior in terms total flying hours and length of time with the company.

This finding suggests that the level of pilot experience influences first officers' ability to accurately identify their own individual performance, given a comprehensive non-technical skills behavioural scale. Therefore, to be required by the company to regularly reflect on their performance on the flight deck might help to facilitate first officer professional development.

One reason for extending the MOSA research to the Indian Subcontinent was to test the methodology in that culture. However, National culture does not appear to have influenced the results as the observed behaviours were very similar to those from the European MOSA study. The present study was conducted in a small airline operating turbo prop aircraft, flying short sectors between island landing strips, although, this type of operational culture also does not appear to have impacted the results.

This study lends support to the body of MOSA empirical research which concludes that non-technical skills self-assessment information collected from professional pilots across a predetermined range of behavioural categories is an accurate indication of performance on the flight deck. Therefore, it is suggested that aggregated and structured aircrew self-reported performance and safety information can be utilized with confidence by management to highlight developing safety issues, and indicate areas of deficiency, as well as identifying the behaviours that work well. Additionally, it has been identified that MOSA methodology can be used in conjunction with other information gathering and evaluation tools to contribute to the on-going safety feedback loop of an organisation's safety management system.

Acknowledgements

The author would like to acknowledge assistance from the following organizations in the conduct of the MOSA studies: the Royal Australian Air Force, Airbus Industries, Easyjet, and Maldivian Airlines.

References

Burdekin, S. (2012) Evaluating Professional Aircrew Safety Related Behaviour In-flight. *In proceedings of the 4th International Conference on Applied Human Factors and Ergonomics* (pp 450–461). San Francisco, USA.

Flin, R., O'Connor, P. & Crichton, M. (2008) *Safety at the Sharp End: Training non technical skills.* Ashgate, Aldershot.

Klinect, J.R., Murray, P., Merritt, A. & Helmreich, R. (2003) Line Operations Safety Audit (LOSA): Definition and operating characteristics. *In Proceedings of the 12th International Symposium on Aviation Psychology* (pp.663-669). Dayton, OH. Ohio State University.

Rural Affairs and Transport References Committee, (2011) *Pilot Training and Airline Safety*. Senate Printing Unit Commonwealth of Australia. Canberra.

United States Congress. (2010) *Airline Safety and Federal Aviation Administration Extension Act of 2010*. United States Government Printing Office. Washington DC

THE NATIONAL AVIATION OPERATIONAL MONITORING SERVICE: DEVELOPMENT OF A SURVEY METHODOLOGY

Mary M. Connors
NASA Ames Research Center
Moffett Field, CA
Robert Mauro
Decision Research
Eugene, OR
Irving C. Statler
NASA Ames Research Center
Moffett Field, CA

The National Aviation Operational Monitoring Service (NAOMS) was a research project designed to develop a methodology for acquiring information on changes over time in safety-related events occurring in the National Airspace System. A scientifically designed survey was used to measure the experiences of front-line operators and to evaluate trends that could compromise safety. Information from NAOMS could be the first indication of a developing situation, providing the basis for further investigation using other sources. This paper reports on a demonstration of the NAOMS concept through development and conduct of a survey of air carrier pilots. Data from over 18,000 randomly selected pilots were taken over a three-year period. Results of this demonstration provide evidence that the NAOMS survey methodology can reliably identify changes over time in the rates of occurrence of safety-related events.

The National Airspace System (NAS) is an exceptionally safe system. However, it is constantly buffeted by technological, procedural, and other changes which may affect the safety of the system. Implementation of the Next Generation Air Transportation System (NextGen, FAA, 2012) will significantly increase the rate at which these changes occur. Several approaches are currently available to assess the impact on the NAS of technological and other innovations. One approach is to compile voluntary reports of safety-related events. This approach, exemplified by the Aviation Safety Reporting System (ASRS), is able to obtain data from across the system. However, because potential respondents choose what and when to report, the accumulated data may not reflect the operation of the system as a whole. Another approach is to use automatically recorded flight data. This approach, exemplified by the Flight Operational Quality Assurance (FOQA) program, can provide statistically reliable data but only some factors can be recorded and only from some users of the airspace system. Several organizations, (White House Commission on Aviation Safety and Security, 1998; Government Accounting Office, 2000; National Transportation Safety Board, 2000,) recognized the gaps in information and called for better data to effectively monitor the status of the NAS. The National Aviation Operational Monitoring Service (NAOMS) project (conducted during the years 2000 to 2005) was established to determine if this need could be addressed through a comprehensive statistically sound survey of NAS operations.

The central purpose of NAOMS was to develop the methodology for a reliable, comprehensive and coherent evaluation of the safety-related experiences of operators of the airspace system – pilots, controllers, maintenance personnel, and cabin attendants. It was hypothesized that a well-designed survey could provide the needed data by relating operator experiences to their corresponding exposure opportunity. This paper briefly describes the NAOMS concept and summarizes the methods developed, the design of the survey instrument for air carrier pilots, the conduct of the air carrier pilot survey, and the results of the analyses of the data obtained. For a more complete report of the NAOMS project, see Connors, Mauro, and Statler (2012)

Concept

The primary goals of NAOMS was to: 1) determine if survey data could be collected in a systematic and objective manner with sufficient statistical power to allow meaningful interpretation, 2) determine the limitations of this approach, and 3) lay the foundation for a future continuing monitoring service. NAOMS was designed to be a reliable pointer to potential safety events that could be further evaluated through other means; it was not designed to provide explanations for these events or to identify appropriate interventions. The NAOMS concept envisioned collecting data from various operators – pilots, air traffic controllers, mechanics, and cabin crews. For this research effort, one group – air carrier pilots – was selected to assess the feasibility of the method. However, the concept could be applied to other operator groups, though the details of the application would need to be specific to each group.

Methods

Because the primary purpose of NAOMS was to determine the feasibility of using a statistically sound survey to obtain information that was unavailable through other means, substantial effort was devoted to resolving methodological issues. The key issues that had to be resolved before the survey could be conducted are described briefly here.

Questionnaire Design

The NAOMS team obtained external reviews of the concept and the desired approach during focus group sessions with air-carrier pilots, consultations with aviation safety and survey methodology experts, and multiple briefings with peers in the FAA and NASA. The opinions of the aviation community were elicited on the events to be addressed, question formats for each event, the order of the questions, and the structure of the survey instrument. Emphasis was placed on identifying the safety topics appropriate for a longitudinal survey designed to reveal changes over time in the rates of potentially threatening events. In its final form, the survey was divided into four sections: Section A obtained information on the respondent's flying during the reporting period. This included information on the hours and legs flown by make and model of aircraft, type of operation – cargo or passenger, and crew position. Flight hours and legs flown are measures of risk exposure; i.e., the opportunity the pilot had to experience a safety event during the time investigated. This "risk opportunity" formed the denominator of the event rates used in the analyses. In addition, some background information (e.g., the interviewee's life-time total commercial flight hours) was obtained. Section B consisted of 96 safety-related questions that elicited information on the number of events experienced by the interviewee during the reporting period. This "risk experience" formed the numerator of the event rates used in the analyses. Section C was included to provide quick looks at topics of interest to the aviation community. Section D asked the respondent about the questionnaire and the survey process.

Survey Mode

To determine the most appropriate way of interacting with respondents, the NAOMS team initially compared in-person, telephone, and mail modes of conducting the survey. Later, the use of an Internet-based survey was explored. It was concluded that telephone interviews represented the best compromise between achieving high response rates and keeping costs within project limits.

Recall Period

One of the most important methodological decisions dealt with the recall period, i.e. the period for which the pilot respondents would be asked to recall safety events. Based on previous research, it was clear that the recall period would affect the accuracy of the recall -- the shorter the recall period, the more accurate the information recalled. However, the longer the recall period, the greater the number of events that are likely to be captured in a single interview. Determining the best compromise between accuracy and the amount of information obtained required substantial investigation. This issue was studied through experiments and small-scale field tests of prototype surveys. The NAOMS team compromised on a 60-day recall period. The general form of the questions posed to the pilot subjects was "How many times during the last 60 days have you experienced [insert specific event.]"

Sample Composition

Obtaining an appropriate pool from which to sample air carrier pilots was a significant challenge. Eventually, it was decided to use the public FAA Airmen Certification database and to randomly select from those United States based pilots who possessed an airline transport pilot certificate, a current first-class medical certificate, a multi-engine rating, and a flight engineer certificate.

Sample Size

Some events happen very rarely and achieving statistical reliability for these events would require an investigation that far exceeded the limits of the NAOMS project. Estimates indicated that 8,000 interviews per year could detect 20% rate shifts with 95% confidence for about half the questions of Section B (Battelle, 2007). The NAOMS team determined that this would be adequate to assess the applicability of the methodology and established 8,000 interviews per year as a goal

Random vs. Panel Sampling

It was decided relatively early in the design process that direct random sampling (without replacement for one year) offered more advantages than did panel sampling.

Interview Procedure

Procedures recommended by Dillman (1978) were followed to engage the pilots and increase participation. Experienced, professionally trained interviewers conducted all interviews. The combination of the factors employed to identify, contact, and engage the respondents resulted in an 81% response rate for those identified and located.

Analyses and Results

The NAOMS air carrier interviews were conducted evenly from January 2002 through December 2004. Data from 18,377 air carrier pilots who met the inclusion criteria were analyzed. (See Connors, Mauro and Statler (2012) for a full description of the limits for inclusion.) Of the 96 questions in Section B, 43 questions captured sufficient numbers of events to allow reliable statistical analysis. The remaining events were too infrequent to allow reliable analysis over the three years of the study. If the study were conducted over a longer period, it is likely that additional questions could be reliably analyzed.

To assess changes over time for the 36-month data collection period, data were grouped into 3 years with 4 quarters in each year. The data were then analyzed to detect linear changes over these 12 quarters. In additional analyses, year-to-year changes and seasonal effects were examined. Because the type of operation can also affect the likelihood of encountering safety-related events, the effects on event rates of aircraft category and cargo/passenger operation also were examined alone and in conjunction with the changes over time.

A series of Negative Binomial regressions were conducted to detect the effects of the predictors described above. Of the 43 questions analyzed in the air-carrier study, 16 questions showed linear trends over the 12 quarters. Fourteen of these events showed reductions in event rates over the time investigated, two showed increasing rates of events. These results indicate that, during this time period, when linear trends were observed, they were predominantly in the direction of greater safety.

Seventeen events revealed significant year-to-year changes. For most of these events, the changes mirrored the linear trends by quarters. However, for three events for which there were no significant linear trends across quarters, there were significant year-to-year non-linear effects.

Statistically significant seasonal effects were observed for 21 of the 43 questions analyzed. The time period represented by each quarter approximated a season of the year, providing an opportunity to assess factors related to seasonal weather (e.g., icing, thunderstorms) or other factors (e.g., tourist travel) that vary regularly across the year.

The operation in which an aircraft is engaged is a major factor affecting the number of specific events encountered and whether some events are experienced at all. The effects of differences among aircraft categories are complex, reflecting the specific types of operations for which these aircraft are used. Significant main effects for aircraft category were found for 36 of the 43 events analyzed. Interactions between aircraft category and year/season were observed for 16 events, indicating that the temporal patterns in these event rates differed by aircraft category.

The survey data were further analyzed according to whether the reporting pilot was engaged in cargo or passenger transport. Of the 43 questions analyzed, four were specific to events involving passengers, and therefore resulted in very few, or no, reports of those events from cargo pilots. For 13 events, there were no differences between operation types. For 20 events, rates were higher in cargo operations than in passenger operations. For six events, rates were higher in passenger operations than in cargo operations.

Full descriptions of the analyses for the 43 questions are reported in Connors, Mauro and Statler (2012).

Considerations for Future Implementations

For any survey to yield valid inferences about a population, it must be based on an appropriate sample. For any method to be practical, it must be cost effective. Before the NAOMS methodology can be implemented on a regular basis, both issues must be addressed for each intended target population.

To obtain a sample of air carrier pilots for the NAOMS survey, pilots were drawn from the public FAA Airmen Certification database. This database does not include employment information. Although all air carrier pilots must have either a Commercial or Airline Transport Pilot certificate, possession of one of these certificates does not guarantee that the pilot is employed by an air carrier. Many pilots obtain these certificates but do not use them. To narrow the pool before attempting to contact potential participants, additional selection criteria were imposed (see Methods above). These requirements resulted in the exclusion of some air carrier pilots. Furthermore, during the NAOMS survey period, the FAA was instructed to allow pilots to "opt-out" of the public Airmen Certification database. This too resulted in the exclusion of some air carrier pilots from the sample. Although there was no reason to expect that the pilots excluded from the sample would experience events at different rates from those included in the sample, analyses were conducted to test this hypothesis. Little evidence of any effect of the exclusions was found. However, similar concerns could arise in selecting samples for other segments (e.g., air traffic controllers, mechanics) of the NAS. In any future implementation of a NAOMS-like system, we expect that all segments of the pilot population would be included and recommend that the most complete databases for every segment be used so that no sample selection issues would arise. Obtaining a simple or stratified random sample would eliminate the need to evaluate the samples for potential biases.

The costs of conducting a telephone-based survey are quite high compared to alternative methods. The NAOMS team recognized that a web-based survey could substantially reduce costs and, near the end of the project, conducted a small study that evaluated this mode. In contrast with the very high response rate obtained in the telephone survey, the web-based approach resulted in a low response rate. However, there may be ways to improve the response rate of the Internet-based survey. It is also possible that a response rate lower than that achieved by the NAOMS telephone-based survey may be adequate. Additional research should be conducted to identify ways in which the integrity and reliability of the survey system can be maintained while lowering costs.

Conclusions

The NAOMS project demonstrated that a scientifically sound survey could provide a statistically reliable method for routinely assessing the status of the National Airspace System along across a range of dimensions. The use of a computer-supported, telephone interview methodology proved highly effective in addressing the main NAOMS objective - identifying changes in event rates over time. It also demonstrated the ability to identify some characteristics of the aircraft and the operations in which the respondents are engaged that are associated with increased vulnerability to encounter specific events. Demonstrated here for air carrier pilots, there is every reason to believe that the NAOMS concept may be applied with similar success to other operational user groups. When combined with information from ASRS, FOQA, and other data sources, a richer, more complete picture of the entire system can be obtained and this knowledge can be used to enhance safety.

Acknowledgements

NAOMS was developed as an element under the Aviation System Monitoring and Modeling (ASMM) project of NASA's Aviation Safety Program (AvSP).

References

Battelle (2007) NAOMS *Reference Report: Concepts, Methods, and Development Roadmap*. NASA Contract NNA05AC07C. November 30, 2007

Connors, M.M., Mauro, R., and Statler I.C. (2012). *The National Aviation Operational Monitoring Service (NAOMS): A Documentation of the Development of a Survey Methodology*. NASA TP-2012-216013, TN #5026.

Dillman, D. A. (1978). *Mail and Telephone surveys: The total design method*. New York: Wiley.

FAA, (2012). NextGen, Implementation Plan, March, 2012.

GAO, (2000). *Aviation Safety*: Safer Skies Initiative Has Taken Initial Steps to Reduce Accident Rates by 2007. GAO/RCED-00-111. June 2000.

NTSB, (2000). *Transportation Safety Databases*. NTSB Report Number: SR--02-02. 9/11/2002

White House Commission on Aviation Safety and Security, (1997). White House Report on Aviation Safety & Security.

ENHANCING HELICOPTER-PILOT OBSTACLE AVOIDANCE USING A BINOCULAR HEAD-MOUNTED DISPLAY

Dennis B. Beringer, Ph.D. and Gena Drechsler
Civil Aerospace Medical Institute
Oklahoma City, OK

Twenty-three helicopter pilots flew a simulated Bell 206 in Helicopter Emergency Medical Service (HEMS) scenarios with obstructions present. In Study 1, the head-mounted display (HMD) showed highway-in-the-sky guidance but not obstructions. Study 2 added obstructions (broadcast towers, power lines) in the HMD. Pilots detected and avoided HMD-depicted obstructions earlier than those only shown out the window. Wire strikes were frequent without HMD depictions of obstructions but were greatly reduced when these objects were shown in the HMD. They were completely eliminated when a red warning fence was overlaid on the power-line graphic at the point it transected the flight path. Pilots indicated that the power-line representation was slightly ambiguous in its meaning, but the red warning fence intent was clear. Pilots preferred power lines without the ground-plane representation and realistic complex green imagery for guy wires, and red/white striped towers with beacons and digital tower elevations.

Helicopter Emergency Medical Services (HEMS) operations, as reported in previous publications (Beringer, Luke, Quate, & Walters, 2009; NTSB, 2006), frequently experience accidents related to obstructions or obstacles to flight that are not as frequently encountered by other types of operations, specifically due to the operation of the aircraft into unimproved and/or confined areas (see Lee, Choi, Choi, & Ujimoto, 2007, and NASDAC, 2002, for accident statistics). As such, parts of the aircraft have been known to strike various structures and both man-made (including other rotorcraft) and naturally occurring objects during the conduct of a flight. One of the major difficulties in these operations, then, is the effective location and avoidance of obstructions; wires of various descriptions are a frequently encountered obstacle. They become particularly problematic in conditions of reduced visibility. Even when visibility is good, descents into confined areas or to unimproved sites are usually very slow requiring multiple crew members to scan the area surrounding the selected touch-down zone for wires and obstructions. That the striking of obstructions by rotorcraft in reduced visibility continues to be a problem is illustrated by the recent fatal crash involving a construction crane atop a high-rise residential complex in London (Topham, 2013).

Means of Detecting Obstructions and Terrain

The use of Helicopter Terrain Awareness and Warning Systems (HTAWS) for rotorcraft engaged in HEMS operations is expected to be mandated soon (Lau, 2013; RTCA, 2008). These systems are anticipated to be coupled with plan-view terrain displays and to provide warnings of approach to terrain and advisories for terrain avoidance (per RTCA DO309 minimum operations performance standards). While this system should allow rotorcraft pilots to detect approaches to terrain and avoid unanticipated terrain contact, it will not necessarily protect against contacting cultural features (human-constructed obstructions).

Some systems that incorporate forward-looking perspective-view displays with some type of terrain representation also include obstruction graphics generated from a database, showing towers and sometimes buildings, usually depicted as some form of vertical lines. However, it may be difficult to ascertain their precise locations in the real world dependent upon the size of and minification in the display. Similarly, obstacles can also appear on plan-view map displays. These systems require the pilot to recognize the location of obstructions by referencing one or more displays and determining a path avoiding the obstacles.

Using head-mounted displays (HMDs) to provide VMC-like visibility of both terrain and cultural features for general aviation applications was outlined in Beringer's (1999a) chapter on this topic. The notion was that one could present terrain, cultural features, and augmented visual information/cues to the pilot to aid in visually guided/referenced maneuvers and piloting in conditions where the prevailing visibility did not allow such with the unaided eye. While the use of head-mounted displays is well established in military combat aircraft, there was not at that time, nor is there now, a certified system that can be used for this type of operation in civil aviation. The existence of efforts to produce a certifiable system for civil aviation made it prudent to evaluate how pilot performance could be affected by the use of such a system, and to what degree both flight performance and thus the level of safety in this type of rotorcraft operations could be improved with such displays.

We previously reported the results of a preliminary evaluation of the use of a HMD for rotorcraft navigation and approach to a landing site (Beringer, et al., 2009) (Study 1) and discussed pilot ratings of various type of graphics that represented obstructions (Beringer & Holcomb, 2010) and the inherent advantages of the HMD. We used a stereoscopic see-through HMD presenting very simple overlaid synthetic imagery of obstructions (Foyle, Ahumada, Larimer, & Sweet, 1992), rather than a synthetic-vision (SV) representation of obstructions (Hughes, 2005; Rash, Russo, Letowski, & Schmeister). This approach was taken to (1) avoid clutter in the visual field and (2) to allow the outside world to be seen through the guidance-information graphics, allowing the pilot to detect any potential hazards when visibility allowed, rather than having them obscured by HMD-displayed data. That work based upon earlier work in a fixed-wing simulator (Beringer et al., 1999b) demonstrated that pilots could effectively use the HMD to follow highway-in-the-sky (HITS) guidance cues to a synthetic landing pad but that they had difficulty using out-the-window obstacle information, as shown by flying through or under power lines on the final approach (as the debriefing discussion frequently went, "Did you see those power-line towers on your final approach?" "Yeah, but I didn't see any wires between them." "Exactly!") and, from the preference assessment in the later study, that they had a definite preference for simple power-line representations and complex broadcast-tower representations.

As is frequently the case in these kinds of evaluations, several questions arose as to the effective use of a HMD with obstructions (non terrain) shown. First, given that the image was derived from position sensing and a database, what would happen if position sensing were a little off or the database contained an error (object in wrong place or object missing altogether)? Would the pilot be able to see through the presented image sufficiently to pick up obstructions that were improperly rendered, or would there be significant enough cognitive or perceptual capture to prevent that? One must add to the "cognitive-capture" category a belief that the image is correct, regardless of information to the contrary, as we have found in other studies of forward-looking terrain displays and NEXRAD displays. Our colleagues in Aircraft Certification frequently refer to this quality in a display as "compelling." Second, would the display of the obstructions alone be sufficient, or would some kind of additional warning be required as a cue to the pilot that the flight path should be altered? That is to say, would our previously preferred low-complexity power-line representations be sufficiently unambiguous to promote avoidance behavior, or would there be enough ambiguity about their location relative to the intended flight path that pilots would not use them effectively? We set out, then, to look at these questions in Study 2 in the same type of applied HEMS scenario as we had before, but added additional graphical imagery to the HMD to evaluate a potential spatial-graphical warning.

Method

Equipment and Displays

Participants flew a simulation of a Bell 206 helicopter with the Chelton/Cobham display suite of a forward-looking perspective primary-flight display (PFD) and a plan-view map/terrain multi-function display (MFD) (Figure 1). Engine instrumentation of a conventional, round-dial B-206 was presented to the left of the PFD and MFD. The Chelton displays contained programmed route information to the remote pick-up site and stick-figure representations of towers. A lateral path vector was also shown for any turning flight and an auditory "obstruction, obstruction" warning was triggered when the path vector came into close proximity to a tower contained in the database, with appropriate visual cuing added to the PFD and MFD.

Participants wore a Kaiser (now Rockwell) Optical stereoscopic full-color see-through HMD (640 by 480 resolution; 25% transmissivity; Figure 1). The display was head-position slaved so that obstructions generated in the HMD overlaid obstructions visible in the out-the-window scene.

Figure 1. *HMD and simulator cockpit with PFD/MFD, engine instrumentation, and out-the-window view*.

Design of stimuli

Six configurations (low, medium, high complexity paired with ground plane symbol off or on) of each HMD obstruction image (radio/television broadcast towers and power transmission towers with lines) from the previous study were used. Red beacons were added to the power-line towers and those on the broadcast towers were expanded vertically to be more visible from a distance. A red-and-white striped broadcast tower similar to real-world towers was added for ratings only. Additionally, one set of obstructions was produced with yellow wires for comparison with green (latter perceived as brighter in HMD, but not a color associated with caution/warning), and a red warning fence was added that could be turned on during the final approach to the helipad. Although all images were rated, the most preferred ones from the previous investigations were used for the flights (low-complexity power lines, complex broadcast towers, green wires). The only change to the synthetic helipad was that the LZ marker was relocated to the top of a pole that started at 200 feet AGL (Figure 2).

Table 1. *Variations in the obstruction images.*

Tower	
Low Complexity (LC)	Vertical line (green), red beacon
Medium Complexity(MC)	Vertical line, guy wires, red beacon
High Complexity (HC)	Vertical rectangles (2)(green), guy wires (green), red beacon; red/white striped version also added
	Ground plane: white circle at radius of guy-wire anchors
Power-line	
Low Complexity (LC)	Vertical line (green), small circle at base (white), line across tops (green), red beacon at top
Medium Complexity (MC)	Vertical lines (2), lines across top (2), red beacon
High Complexity (HC)	Trapezoids (2), splayed, lines across top (2), red beacon
	Ground plane: wire shadows for low and medium; square tower base for high
Overall color	Green guy wires and power lines versus yellow of same
Power line overlay	Red "fence of death" overlaid on power-line segment intersecting intended flight path

In an attempt to conserve space, only the full scene used for the flight trials (Figure 2, left) and the warning graphic for the power line (Figure 2, right) are shown in this paper. See Beringer & Holcomb (2010) for the other variants used during the ratings portion as they were repeated from that study.

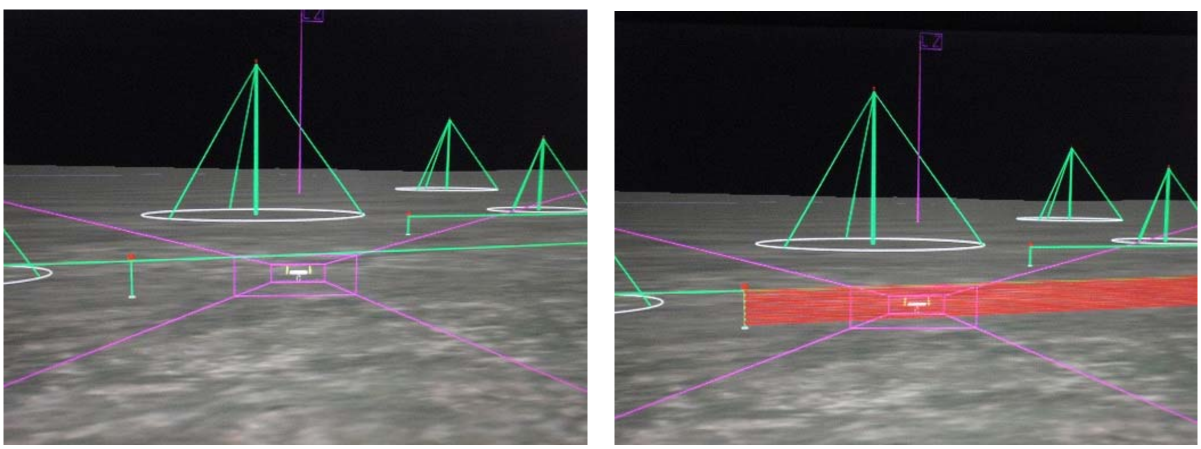


Figure 2. *Obstacle field and HITS (left) compared with same view having red power-line warning fence (right).*

Participants/Procedure

Seventeen licensed rotorcraft pilots from the greater Oklahoma City area participated in Study 2, with 14 completing all flights. Of these, 6 had ATP/helicopter ratings, 6 were 90-day current, and 8 had logged time in the previous 12 months. Data for six pilots from Study 1 where obstructions did not appear in the HMD were included for comparison with the Study 2 trials in which obstructions were shown.

Each participant first took a familiarization flight in the simulator in the vicinity of KAMA (Amarillo) and the mythical HEMS base. After accomplishing this, each was fitted with the HMD and flew a baseline flight from KAMA direct to the remote pick-up site with obstacles represented in the HMD but without any explicit explanation of what they represented. This course line took them very close to a tall broadcast tower near the destination. Following this segment, the simulator was placed in freeze mode in a position where they could view the various obstruction representations in the HMD. Each viewed a complete subset of stimuli (power lines first, then broadcast towers, then green versus yellow obstacle fields, then 4 with striped broadcast towers) without rating, then viewed the same subset again while assigning ratings (5 scale points from poor to adequate to very good) and making unstructured comments about each graphic. This was done for each of the subsets until all graphics had been rated.

Following the ratings, the participants flew 3 more flights using the HMD. Flight scenarios followed those in the 2009 study and used the same type of pathway guidance for the later trials (approach to a remote landing site using a highway-in-the-sky 6-degree descent guidance to a synthetic helipad seen in the HMD with out-the-window VMC conditions; 3-mile visibility in haze). The first of these was a familiarization flight with the HITS guidance flying from the HEMS base at KAMA to KTDW (Tradewinds, west of KAMA) and an actual helipad. Pilots used the PFD/MFD to fly a HITS on the PFD to the intersection with the descent to KTDW, where another HITS was depicted in the HMD (consistent with leg 2 of the HITS on the PFD). This was identical in concept to what would come next, as the next flight was from the HEMS base to the pick-up site NW of KAMA in the midst of HMD-presented broadcast towers and power-line runs, using guidance on the PFD and MFD to intersect the final approach course/6-degree descent to the synthetic helipad. Emergency vehicles and an accident vehicle were present at the site in the out-the-window view.

We saw some of the same behavior (flying through or below wires) in the earliest trials, even though the power lines were explicitly depicted. We added the red warning fence at this point for the remainder of the participants so that we could compare an active indication of a hazard with the extant passive representation. This was added as a fourth flight to all of the later sessions and was activated half way along the final descent, a point before which we would have expected to see any flight-path deviation according to our initial performance data.

Results & Discussion

Three issues will be addressed in this reporting. First, how did the pilot ratings of graphics compare regarding what this sample found most representative of the obstructions that they were intended to represent, and was there a color preference overall? Second, we wanted to assess to what degree having the obstructions visually represented in the HMD would influence course deviation on flights without overt 3-D path guidance and closeness of approach to the obstructions (here we report only the nearest approach to first and tallest tower). Third, we wanted to determine how effective the power-line warning graphic was at causing pilots to appropriately alter their flight path.

Pilot Ratings of Graphics

Rating data for 13 of the pilots were found complete enough to be usable. Participants assigned a score of 1 to 6 to each graphic, 1 representing “very poor” and 6 representing “very good.” Figure 3 shows mean pilot ratings for tower depictions and 1 standard deviation above and below the mean.

Presence of the ground-plane indicator did not greatly affect ratings within any level of complexity, but there was a slight preference for the more complex depictions (guy wires favored). Removing the tower mast was seen as negative, and the threat cylinders were not seen favorably at all. Comparing the most preferred format, striped tower masts having guy wires, ground circles, and digital altitude markers, with the cylinders indicated that the distributions of ratings did differ significantly [$t(12) = -4.15$, $p<.05$], consistent with the negative comments about the cylinders provided by the raters. One can also see a small but consistent increase in ratings of the masts as each additional feature was added (ground-plane circle, guy wires, and digital altitude).

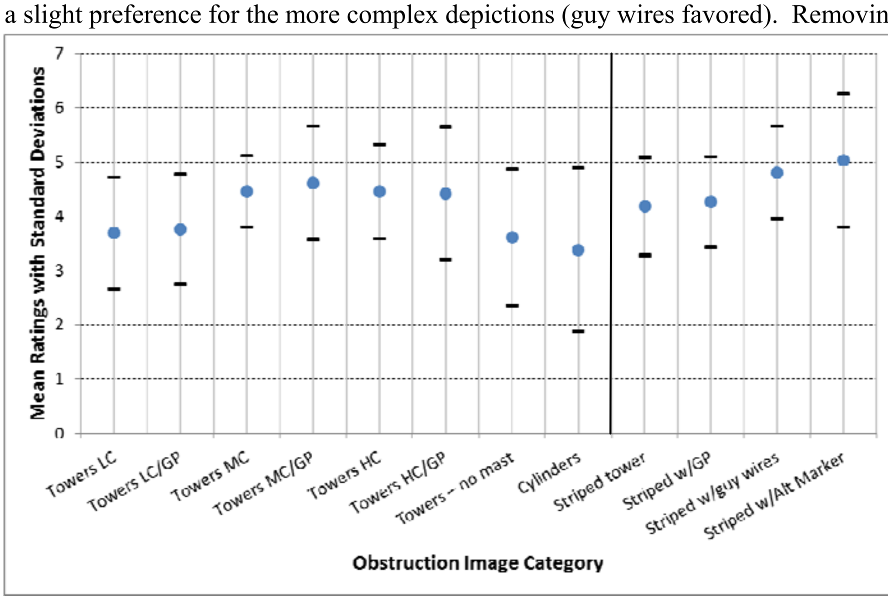


Figure 3. *Means and standard deviations for pilot ratings of broadcast towers.*

Regarding the power-line representations, the differences were not significant excepting between the lowest-rated one (LC with a ground-shadow line) and the first high-complexity tower (HC, no ground plane indicator) that had dual lines on top of paired trapezoids [$t(12) = -2.54$, $p < .05$]. There was uniform agreement that anything producing what looked like a ribbon or chain of boxes looked too much like command guidance information (HITS) and was confusing. Thus, adding the ground-plane was not as advantageous as keeping the depiction simple, particularly as this was only a factor when close to 120 feet AGL. Although a verbal preference for green wires was often expressed, ratings for green and yellow did not differ significantly despite a slightly higher mean for green.

Approach to Tower

Figure 4 shows a plot of the closest approach, on the initial flight to the NW field of obstacles, by participants from both Study 1 and Study 2. Note that points indicated by arrows in Figure 4 were those individuals in Study 1 having no HMD-presented obstructions but having the destination synthetic helipad visible (obstacles only visible in the out-the-window view with towers but no wires).

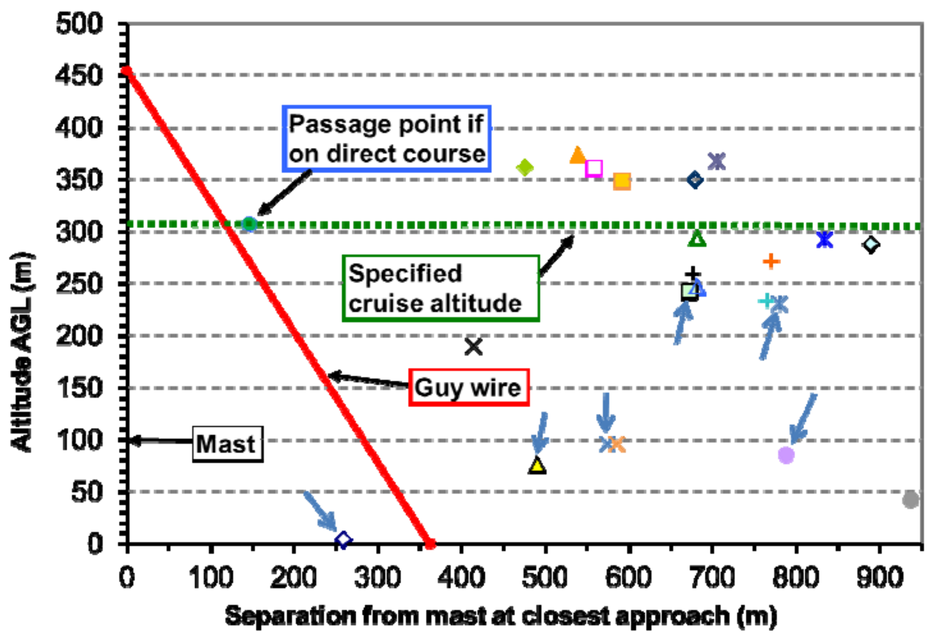


Figure 4. *Altitude and lateral separation from tower mast/guy wire at closest approach.*

The remainder had GPS guidance via the MFD/PFD and obstacles with wires shown in the HMD but no synthetic helipad/destination. From the distribution by groups, it is evident that those without HMD obstructions and with line-of-sight HMD visibility of the destination tended to descend earlier and thus place themselves in potentially a greater hazard regarding tower guy wires.

Point of Course Deflection & Power Line Avoidance

Space constraints prohibit a detailed discussion herein of the course-deflection findings, which will be reported in the technical report version of this paper. However, to roughly summarize, pilots with the obstructions depicted in the HMD (1) reported the obstructions as a potential hazard earlier, (2) appeared to make intentional course deviations away from the direct course earlier, and (3) maintained altitude longer before descending to the destination.

Of the 14 pilots who completed the third and fourth flights using the HITS, 5 flew through power lines on the final approach to the pick-up site on the third flight, replicating our earlier findings to some degree, although with a slightly reduced frequency of occurrence. There was no discussion of the pilots' performances between the third and fourth flights, and they were told only that they would see something different on their displays and that the approach would be slightly different. On the final flight, **none** of the pilots flew through the power lines, with the minimum clearance over the lines being 69 feet at crossing (mean without warning fence, 144 feet; mean with warning fence, 306 feet). We hypothesized that the warning would allow pilots to analyze that the HITS would take them through the lines and thus arrest their descent earlier and allow better vertical separation from the lines (Paired-samples t, one-tailed: $t(13) = -1.851207$, $p<.05$). Pilots indicated, in the post-flight debriefing, that they would like an integrated system in which the MFD, PFD, and HMD indications were all correlated and synchronized so that all warnings regarding obstructions appeared at the same time and in related representations on all 3 forms of display, as well as in the auditory obstruction warnings generated for towers by the MFD/PFD equipment.

Conclusions

Preferences for HMD graphics, as seen in previous assessments, tended towards power-line graphics without the ground-plane shadow but towards the more literal representation (HC) for broadcast towers because the graphic looked like a tower with guy wires. Flight-performance results demonstrated that obstruction imagery in a head-mounted display could be useful in aiding rotorcraft pilots engaged in an HEMS-like activity to avoid obstacles, both strategically and tactically. This was particularly true in a case that simulated a misprogrammed navigation system that would take the descent path through a known obstruction, in which case the inferred conflict (passive obstacle representation) was successful in preventing a collision only 64% of the time. However, the addition of an active graphical warning overlaying the hazard was completely (100%) effective in preventing power-line strikes during the final descent. The possibility of using HMDs for both navigation and obstacle avoidance in operations near the surface of the terrain appears to be supported by the cumulative results to date.

Acknowledgements

Research reported in this paper was conducted under the Flight Deck Program Directive of the Human Factors Division (ANG-C1), FAA Headquarters, and the Aerospace Human Factors Division (AAM-500) of the Civil Aerospace Medical Institute, Federal Aviation Administration.

References

Beringer, D.B. (1999a). Innovative trends in general aviation: Promises and problems. Chapter 10 in David O'Hare (ed.), *Human Performance in General Aviation*. Brookfield, VT: Ashgate Publishing Co., 225-261.

Beringer, D.B. (1999b). Flight command-guidance indicators or pathway displays: My way or the highway? In *Proceedings of the 10th International Symposium on Aviation Psychology*, 57-62.

Beringer, D.B., Luke, T., Quate, A. & Walters, E. (2009). Helicopter pilot use of a see-through, head-mounted display with pathway guidance for visually guided flight: observations of navigation behavior and obstacle avoidance. *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society*, 26-30.

Beringer, D.B. and Holcomb, K. (2010). Factors influencing the effectiveness of simple obstacle-related graphical depictions for head-mounted displays. In *Proceedings of the 54th Annual Meeting of the Human Factors & Ergonomics Society*, 75-79.

Foyle, D.C., Ahumada, A.J., Larimer, J., and Sweet, B.T. (1992). Enhanced/synthetic vision systems: Human factors research and implications for future systems. SAE Transactions: *Journal of Aerospace, 101*, 1734-1741.

Hughes, D. (2005). Virtual VFR is coming. *Aviation Week & Space Technology,* 163(21), 58-59. Retrieved September 22, 2008, from Academic Search Elite database.

Lau, S. (2013). Helo Technology: HTAWS to be mandated on all EMS Helicopters. In *Professional Pilot*, Wednesday, January 30, 2013.

Lee, Y.H., Choi, Y.C., Choi, S.H., & Ujimoto, K.V. (2007). Analysis of survey data on situation awareness of helicopter pilots. Transportation Research Record: *Journal of the Transportation Research Board*, 111-116.

National Aviation Safety Data Analysis Center (2002). NTSB Helicopter Accident Study. Retrieved September 16, 2008, from http://www.nasdac.faa.gov/aviation_studies/ntsb_helicopter _accident _study/helicopter_accident_study.html

National Transportation Safety Board (2006). Special investigation report on emergency medical services operations. Special Investigation Report NTSB/SIR-06/01. See http://www.ntsb.gov/publictn/2006/sir0601.pdf

Rash, C.E., M.B. Russo, T.R. Letowski, & E.T. Schmeisser (Eds.). (n.d.). Helmet-mounted displays: Sensation, perception and cognition issues. Fort Rucker, Alabama: U.S. Army Aeromedical Research Laboratory, ISBN 978-0-615-28375-3. (no publication date in text)

RTCA (2008). Minimum operational performance standards (MOPS) for helicopter terrain and warning systems (HTAWS) airborne equipment. RTCA/DO-309. Washington, DC: RTCA, Inc.

Topham, G. (2013). London helicopter crash pilot decided to pick up client despite warnings. In *The Guardian*, Wednesday, 23 January, 2013.

AN INVESTIGATION AND ANALYSIS OF THE VESTIBULO-OCULAR REFLEX IN A VIBRATION ENVIRONMENT

Daniel Uribe and Michael E. Miller
Air Force Institute of Technology
Wright-Patterson AFB, Ohio
Suzanne Smith
711th Human Performance Wing
Wright-Patterson AFB, Ohio

Helmet Mounted Displays (HMDs) and their integration into military systems have greatly improved. However, previous research has demonstrated degraded visual performance when an HMD user is subject to whole-body, low-frequency vibration. This effect has been attributed to the effect of the Vesitbular-Ocular Reflex as it stabilizes the eye with respect to the external environment, causing eye movement with respect to the HMD. This research sought to understand the VOR as a function of whole-body, low-frequency, z-axis vibration. A human subject experiment was executed to measure the effect of whole-body, low-frequency vibration on eye movements recorded with Electro-oculagraphy (EOG) while performing visual fixation tasks on an HMD. The results indicate that during fixation on a stationary target, the magnitude of VOR-driven eye movement was greatest for a vibration frequency range of 4-6 Hz. The findings are consistent with previous research in visual performance.

The Head- or Helmet-Mounted Display (HMD) continues to advance technologically, but also in its applications. For example, F-35 pilots are highly dependent on their HMD as it has replaced the heads-up and panel displays employed in previous military aircraft. By eliminating practically all other in-aircraft displays, the HMD reduces aircraft weight and power, while providing novel capabilities, such as the ability to "look through" the hull of the aircraft below the canopy. HMDs have also become a necessary component of the human interface within many rotor-wing platforms and are being used by dismounted personnel and personnel driving land based vehicles. HMDs are popular in these applications as they provide critical information with increased mobility (Geiselman and Havig, 2010). The fact that HMD permits the designer to display information within the user's visual field, regardless of head location, improves performance by enhancing situational awareness (Velger, 1998).

While HMDs have provided positive improvements in aircrew and battlefield performance (Daetz, 2000), many significant issues remain with this technology, such as visual clutter, attentional capture, and inattention blindness that can occur from using bright, collimated displays, such as HMDs (Gibb et al., 2010). Another significant problem occurs when these displays are viewed by an operator in an environment that is undergoing low-frequency vibration. Under these conditions the effectiveness of the HMD decreases sharply due to a loss in perceived resolution that is thought to occur as a result of the human Vestibular Ocular Reflex (VOR) (Rash, et al, 2009). While most fixed-wing, jet-propelled aircraft are not subject to these low-frequency vibration regimes for significant lengths of time, propeller driven aircraft occupants, helicopter operators, and land vehicle drivers can routinely experience these vibration ranges throughout a mission. Further, certain fixed-wing jet aircraft are subject to buffeting which exposes the pilot to similar vibration, although for relatively short intervals of time (Daetz, 2000).

While previous research has demonstrated reduction in human visual performance when using HMDs in environments undergoing vibration, these studies have generally not sought to understand the relationship between eye and head movement within these environments. Therefore, the focus of this research was to understand the motion of the eye and head as the user performs simple visual tasks. It is hoped that this information can aid the design of future vibration compensation systems.

Background

Vibration occurs in six-degrees of freedom, e.g., x, y and z dimensions, as well as roll, pitch and yaw. Vibration is highly dependent on many different variables, including aircraft type, seat type, body size, muscle tone, posture and helmet weight (Daetz, 2000). The human body is physically sensitive to vibration. Numerous studies have established that whole-body resonance occurs within the frequency range between 4 and 8 Hz during exposure to vertical vibration. Head pitch magnitude occurs most severely between 4 and 8 Hz as a result of the forces that occur during resonance (Smith, et al, 2007). Head translational and rotational transmission is generally greatest between 0 and 10 Hz, while most vibration occurring at frequencies greater than whole-body resonance, e.g., greater than 10 Hz, is typically dampened before reaching the head (Paddan and Griffin, 1988).

Vibration has been shown to degrade human visual performance within a number of studies. Lewis and Griffin demonstrated that reading performance was degraded most severely for frequencies between 5 and 11 Hz, resulting in reading error rates up to 20% for subjects seated in a vibrating helicopter seat (Lewis & Griffin, 1980). A separate study, in which participants were subject to display-only, participant-only, and both display-participant vibration, found that the reading performance of the display-only vibration test rated significantly worse than the other two (Moseley and Griffin, 1987). These vibration effects also extended to more complex tasks associated with military operations. In general, the highest or most severe visual performance degradation occurs at frequencies in which vibration is transmitted to the head. While various research studies have demonstrated that the frequency range of the most severe visual performance degradation is between 4 and 6 Hz, at a magnitude of 1 m/s^2 or .1 g, this research has also shown that visual performance can be adversely impacted at higher frequencies up to 20 Hz (Rash, 2009; Velger, 1998; Griffin, 1990). In addition, high intensity vibration beyond 20 Hz that does reach the head can cause visual blurring due to resonance behavior in the eye or associated structures (Griffin, 1990).

The human eye generally moves in fast saccadic movements between relatively stable fixation points. Once fixated on a point, the human eye then integrates information from the scene. An alternate eye movement behavior occurs when a user is tracking a moving object. During these events, the eye moves in smooth pursuit, often referred to as fixation reflex during which the eye is fixated on an object of interest and follows this object in space, typically without the need for head movement. However, for fast moving objects, smooth pursuit eye movements can be coordinated with slow head movements to permit the object to be tracked over large distances. Importantly for either type of eye movement, the eye time-integrates the information from the scene, permitting the object of interest to be imaged onto the retina and captured at a high signal to noise ratio.

The degradation of visual acuity and thus task performance due to vibration when using an HMD has been attributed to the eye and the associated motions induced by the VOR within vibratory environments. When rotational (roll, pitch and yaw) head movements, such as those caused by vibration, are introduced to the head, the eye responds with a reaction known as VOR. The VOR occurs when the semicircular canals in the ear detect head motion, sending a signal to the eye muscle, which induces the eye to involuntarily move in the direction opposite to and near the same magnitude as the movement of the head (Tabak, et al, 1997). Some studies have found the VOR to be effective only at frequencies up to 10 Hz, though other research suggests that the VOR is operative in frequency regimes of 20 Hz and higher (Griffin, 1990). For example one study found that the errors associated with the tracking of a target with both the eyes and the head increased nearly linearly with increased vibration amplitude for frequencies between 2 and 20 Hz (Shoenberger, 1972).

The VOR allows whatever image is being focused on, to remain in the center of the retina while the head undergoes motion, facilitating the integration of information from the image. This reflex is most noticeable when doing high-impact activities such as running or jumping. During these activities, the VOR allows the world around us to be stabilized in space by moving our eyes to adjust for head motion and allowing us to fixate on objects within our environment. However, the VOR is not effective in the presence of translational head motion, because this head motion causes an angular displacement of the image, which is dependent upon the distance of the object from the eye, which the VOR cannot correct (Griffin, 1990). The VOR assumes that our natural world is stationary and we move within the world reference frame. Therefore, when visual information is provided on an HMD which moves with the head, this VOR response becomes inappropriate as the image provided by the HMD is stabilized with respect to the head. Since the VOR causes the eye to move to compensate for head movement, this compensatory eye movement results in relative motion of the display with respect to the eye.

Another complicating factor is that it cannot be assumed that the VOR correctly compensates for head motion under all circumstances. The pursuit, or fixation reflex, allows the eye to follow the data or text at vibrations at around 1 Hz, but at frequencies higher than 3 Hz, reading the vibrating display isn't only difficult, it becomes nearly impossible as the fixation reflex can no longer keep up with the movement caused by vibration. (Griffin, 1990) Therefore, it is believed that the VOR induces relative motion between the HMD and the human retina that is not predictable. Unfortunately, while the VOR is an essential part of our human composition, when viewing HMD's while seated in environments undergoing low frequency vibration, e.g., less than 10 Hz, the VOR can cause significant degradation in performance and this degradation may not be predicted by measuring the motion of the user's head alone. Further, the head is not likely to vibrate synchronously with the environment as the human skeleton and tissue is likely to affect the transmission of vibration from an environment to the user's head.

While vibration and the associated VOR effect are well documented, little has been done in the past 40 years to compensate for this effect. The rudimentary solution has been to simply increase the size of the text or graphics (Griffin & Lewis, 1978), space out the text (Griffin, et al., 1986) or change contrast levels (Moseley &

Griffin, 1987) on the display. These adjustments result in improved reading performance under vibration conditions. Unfortunately, these recommendations run counter to HMD technology investment to further the resolution of HMDs in an effort to increase the rate of information transfer between the system and the human operator.

Another form of compensation involves image stabilization, by attempting to move the displayed information synchronously with the motion of the head to compensate for the VOR. Many compensation methods have been attempted, including measuring acceleration signals from the helmet and applying a double-integration filter to displacement and amplifying this signal to the HMD electronics to adjust the image accordingly (Wells and Griffin, 1984). Another compensation method applies adaptive noise cancellation to create a compensation input for an HMD. This type of compensation takes into account the reference of a primary signal, often from an aircraft mounted accelerometer which measures vibration levels, and then filters it through a biodynamic transfer function, which estimates the seat to head vibration transmission based on the aircraft type and flight regime. This filtered signal then moves information on the HMD synchronously with the sensed aircraft vibration to provide image stabilization for viewing tasks (Lifshitz and Merhav, 1991; Velger, 1998). A third method involves applying a conventional notch/lag filter to remove the effect of head rotation (Daetz, 2000). While each of these methods found varying degrees of success, the resulting algorithms suffered from inaccuracy and latency issues. Because VOR-induced eye movement was not understood, it is unclear whether the error is due primarily to latency or whether our lack of understanding of the VOR when viewing an HMD produced unexpected errors.

Improving the accuracy of compensation algorithms requires a more in-depth understanding of the VOR. VOR research has been conducted in the medical field with a focus on determining patient vestibular deficiencies. Two such studies were conducted by applying a helmet apparatus to perturb the subject's head in the low-frequency domain. One study analyzed the VOR effect based on fixation of a stationary target, while later research incorporated head-free tracking of a moving visual target. Both studies found that the VOR was predictable and acted linearly up to approximately 4 Hz. However, they found that the dynamics of the VOR began to vary at greater than 4 Hz (Tabak, et al., 1997; Tangorra, et al., 2004). The Tabak study, which involved fixation on a target, found that VOR gain *decreased* up to 8 Hz, and then increased. The Tangorra study, which involved tracking a moving target, found the VOR gain *increased* at frequencies greater than 4 Hz. Both studies concluded that the VOR has non-linear dynamics at higher frequencies and suggested the need for further evaluation. These studies did attempt to control the vibration of the head by using a head perturbation system and did not induce whole-body vibration.

The present study, conducted at vibration frequencies encountered by rotorcraft and during buffeting, sought to validate these findings and provide a foundation to further research on compensation algorithms.

Methods

Participants

Six volunteers between the ages of 20 and 26 years with a mean of 23 years participated in the study. Participants had not experienced any vestibular anomalies, including inner ear infections, within the month preceding the investigation or reported any discomfort or pain symptoms associated with the musculoskeletal system. Additionally, female participants were not pregnant and did not have breast implants. Individuals requiring glasses or hard contacts were precluded from participating in the experiment.

Apparatus

During the experiment, participants wore an HMD system, which supported visual tasks. A custom HMD was designed and built to accomplish this goal. This system included a typical Air Force Flight Helmet, equipped with a visor. A binocular display was mounted on the visor, which permitted a VGA image to be provided to each eye with a field of view of 30 degrees. This helmet supported the six electrodes, attached to the user's face, used to measure the movement of the eyes with respect to the head via EOG. EOG measurements were facilitated using a BioPac MP150 which permitted EOG signals to be obtained at a sample rate of 1000 Hz. This procedure recorded the potential difference between the electrodes as the eye moved from the center, neutral position towards the electrodes. Additionally, the BioPac system was equipped with a 3 DOF accelerometer, which was attached to a custom fit mouth guard for each participant to permit head acceleration to be measured.

This study was conducted in the Single-Axis Servo-hydraulic Vibration Facility supported by the Air Force Research Laboratory's 711th Human Performance Wing. The human-rated single-axis vibration table is capable of generating various vibration signals in the vertical or Z direction. A rigid seat with seat pan and seat back cushions was mounted on top of the platform. For this study, single sinusoidal frequencies were generated between 0 and 10 Hz in 2 Hz increments with an amplitude of 0.1 g Peak.

Experimental Procedure

During the experiment, participants completed two tasks. Task A required the participant to fixate on a centered, stationary target, while Task B required the participant to fixate on a moving target which moved around the display in a rectilinear pattern. This target increased in velocity from each apex in the pattern, reaching a rate of 190 pixels per second and maintaining this rate before decelerating as it approached the next apex. Participants completed two trials, performing each task at each vibration frequency on two separate days.

The participants were exposed to a vibration condition for 20 seconds and then performed the task for 15 seconds. Once completed, the frequency of vibration was changed and the task repeated for the subsequent vibration condition. The task order, as well as frequency exposure order, was varied over the two trials.

Data Analysis

The resultant EOG recording data were processed and analyzed using code developed in MATLAB. Independent variables for this research included the vibration frequency and trial. The primary dependent variable of interest was the magnitude of vertical eye movement, indicated by a change in the EOG signal measured in mV.

The vertical component of the EOG data was processed using the following steps. First, the signal collected during the fixation in the first task or horizontal target tracking in the second task was segmented from the data for further analysis. The segmented data was first filtered to eliminate low-frequency drift in the EOG signal by applying a low-pass filter. A second filter was applied to determine high amplitude values caused by blinks or other erroneous eye movements and this data was segmented out of the data stream. Finally, the remaining data were analyzed to determine the Root Mean Square (RMS) value in mV for each participant at each frequency condition during each trial. The RMS was calculated using Welch's method (Diez, 2008), which estimates the power spectral density estimate (PSD) at different frequencies using an overlapping window principle and computing the discrete Fourier Transform. The resulting array was the RMS value as a function of frequency for each of the two tasks. The peak frequency was then selected and was found to consistently correspond to the frequency of the input signal for all frequencies between 2 and 10 Hz.

Results and Discussion

An analysis of variance (ANOVA) was applied to both of the individual tasks to understand whether there was a statistical effect of Frequency or Trial on the RMS of the EOG signal. Participants were treated as a random effect within this model. The ANOVA for the stationary fixation found that Frequency had a significant effect on the eye magnitude value ($F(4, 20.76) = 23.29$, $p<0.0001$). To further understand this interaction, a post hoc Tukey test was conducted on the 5 vibration frequency conditions to determine statistically significant differences between the levels. The Tukey test used applied a 95% confidence interval ($\alpha = .05$) and calculated the Least Square Values (LSM) of the data at each frequency. The mean RMS value at 4 Hz was significantly higher than the mean RMS values for the remaining frequencies, while the mean RMS value for the 6 Hz frequency condition was significantly higher than the mean RMS value for the 8 and 10 Hz conditions. Figure 1 shows that the mean RMS values increase dramatically as the frequency is increased from 2 to 4 Hz and then decreases with further increases in frequency.

The same ANOVA model was applied to Task B, the moving fixation task. This analysis found that, as expected, Frequency had a significant effect on the eye magnitude value ($F(4, 20.93) = 22.57$, $p<0.0001$). A post hoc Tukey test was again conducted to determine statistically significant differences between the vibration conditions. As in Task A, the results for Task B found that the mean RMS value for the 4 Hz condition was significantly higher than all other conditions, while the mean RMS value for the 6 Hz condition was significantly higher than the mean RMS values for the 4 and 10 Hz conditions. Figure 2 shows that, as in the single point fixation task, the magnitude of the mean RMS value increases to a peak at 4 Hz and declines for higher frequencies.

Previous research regarding the impact of the VOR on visual performance in vibration conditions had relied only on performance measurements, such as reading or aiming errors. This research had indicated an increase in aiming error and acuity decrements (Velger, 1998), as well as a rise in reading errors (Wells and Griffin, 1990) within the 4 to 6 Hz range. Additionally, the literature suggested that as the vibration frequency increased beyond 6 Hz, the body would dampen the vibration and decrease the transfer of vibration to the head (Rash, 2009). The results of the present study confirmed that the highest VOR-induced eye movements did in fact occur at this 4-6 Hz range, while steadily decreasing with increasing frequency. Overall, the findings for analysis performed on each task found that the RMS amplitude of the vertical eye movements as a function of frequency followed the same general pattern

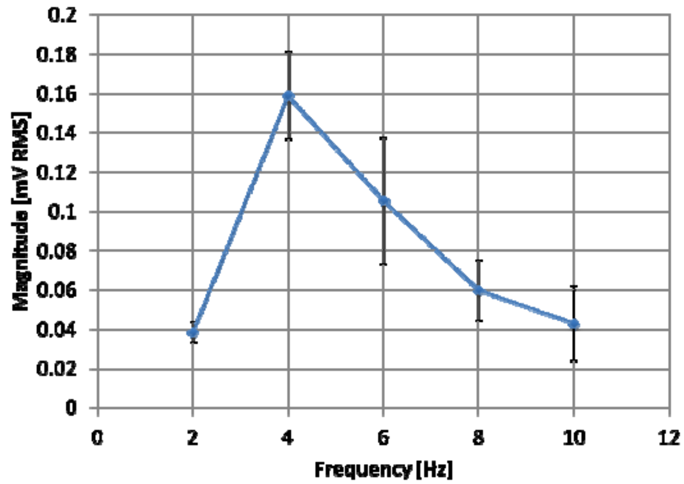


Figure 1. Mean RMS value (in mV) as a function of vibration frequency, with one standard error for Task A

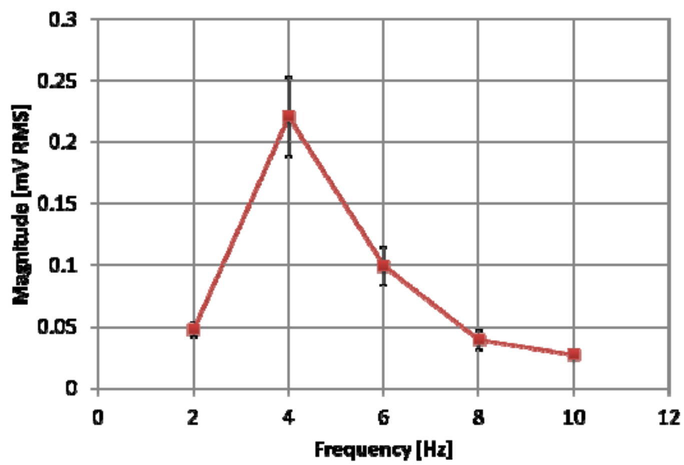


Figure 2. Mean RMS value (in mV) as a function of frequency, with one standard error for Task B

Conclusions

This investigation delved into a research area that until now has been largely overlooked, namely investigation of the compensatory eye movement of the VOR for an individual undergoing whole-body vibration while wearing an HMD. This research allowed a participant to perform visual tasks on an HMD, while also capturing the participant's eye movement via EOG. The results from fixation and tracking tasks corroborated with decades of visual performance research, finding that the VOR-induced eye movement as a function of vibration frequency has a maximum magnitude in the 4-6 Hz range. This study gathered data over only a small sample of participants and does not lend itself to understanding individual differences in eye movements as a function of frequency. However, this investigative effort was able to show that eye movement data can be collected in a vibratory environment and that it provides information that is consistent with past performance research. Further analysis of this data should focus on a comparison of the tasks to determine an effect, if any, of adding motion to a task. Additionally, the data measuring head acceleration needs to be examined in an attempt to better understand the relationship between head and eye movement in an HMD application.

Acknowledgements

The authors would like to thank the Faculty Research Council at the Air Force Institute of Technology for financial support and the Navy Aeromedical Research Unit in Dayton for support of EOG measurement.

References

Daetz, D. (2000). *Development of a Biodynamic Interference Suppression Algorithm for a Helmet Mounted Display Tracking Task in the Presence of Aircraft Buffet*. MS Thesis. School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB, OH, 2000.

Diez, P. (2008). "A comparative study of the performance of different spectral estimation methods for classification of mental tasks," *Engineering in Medicine and Biology Society, 2008. 30th Annual International Conference of the IEEE, 20-25 Aug. 2008*

Geiselman, E. and P. Havig (2011). "Rise of the HMD: the need to review our human factors guidelines", *Proceedings of SPIE 8041,* Head and Helmet-Mounted Displays XVI: Design and Applications, 10 May 2011.

Gibb, R., R. Gray and L. Scharff (2010). *Aviation Visual Perception: Research, Misperception and Mishaps*. Farnham, Surrey, England: Ashgate, 2010

Griffin, M. (1990). *Handbook of Human Vibration*. London: Academic, 1990

Griffin, M and C. Lewis (1978). "A review of the effects of vibration on visual acuity and manual control, Part I: *Visual Acuity"*, 1978, 383-413

Griffin, M., R. McLeod., M. Moseley, and C. Lewis (1986). "Whole body Vibration and Aircrew Performance: *ISVR Technical Report 132*."

Lewis, C. and M. Griffin (1980). "Predicting the effects of vibration frequency, and axis and seating conditions on the reading of numeric displays," *Journal of Sound and Vibration,* Volume 70, 3, 355-377

Moseley, M. and M. Griffin (1987). "Whole-body vibration and visual performance: An examination of spatial filtering and time-dependency," *Ergonomics*, 613-626.

Paddan, G. and M. Griffin (1988). "The transmission of translational seat vibration to the head: Vertical seat vibration," *Journal of Biomechanics,* Volume 21, 3, 191-197.

Rash, C., et al. (2009). *Helmet Mounted Displays: Sensation, Perception and Cognition Issues.* Fort Rucker, AL.: U.S. Army Aeromedical Research Laboratory, 2009.

Shoenberger, R. (1972). "Human Response to Whole Body Vibration," *Perceptual and Motor Skills,* Volume 34, 127-160.

Smith, S., et al. (2007). *Head transmissibility characteristics during single and combined-axis vibration exposures.* Proceedings of the 42nd United Kingdom Conference on Human Responses to Vibration, Institute of Sound and Vibration Research, University of Southampton, Southampton, England, 10-12 Sep.

Tabak, S., et al. (1997). "Gain and delay of human vestibulo-ocular reflexes due to oscillations and steps of the head by a reactive torque helmet," *Acta Oto-laryngologica*, Volume 117, 6, 785-795.

Tangorra, J., L. Jones and I. Hunter (2004). "System identification of the human vestibulo-ocular reflex during head-free tracking," Bioinstrumentation Laboratory, Department of Mechanical Engineering, MIT, Cambridge, MA.

Velger, M. (1998). *Helmet Mounted Displays and Sights*. Boston: Artech House, 1998.

VISUAL PERSPECTIVE ILLUSIONS AS AVIATION MISHAP CAUSAL FACTORS

Frederick R. Patterson, Ph.D.
Naval Medical Research Unit Dayton
Wright-Patterson Air Force Base, Dayton, Ohio

Richard D. Arnold, Ph.D.
Naval Medical Research Unit Dayton
Wright-Patterson Air Force Base, Dayton, Ohio

Henry P. Williams, Ph.D.
Naval Medical Research Unit Dayton
Wright-Patterson Air Force Base, Dayton, Ohio

In the past, aviation spatial disorientation (SD) has been considered predominantly an isolated vestibular problem, associated with lack of outside visual cues. Recent research has challenged the prevailing position by suggesting pilot SD is more commonly caused by problems with cognitive processing of visual spatial strategies. Among the confounding visual stimuli known to occur in-flight, several visual perspective illusions have been identified as reoccurring mishap causal factors. These illusions occur in part because humans exhibit a strong instinctive tendency toward considering themselves level with distant horizontal references, regardless of their true altitude. Also contributing to perspective illusion occurrence is the fact that objects viewed slightly above the perceived horizon will appear magnified in size and exaggerated in elevation above actual aircraft altitudes. Although perspective illusions have been recognized and studied for centuries, the impact of these sensory misperceptions on aviation safety has only recently been subjected to close scientific scrutiny.

Aviation spatial disorientation (SD) is best described as a pilot's inability to correctly interpret aircraft attitude, altitude, or airspeed in relation to the earth or other points of reference. If not recognized immediately, this sensory misperception can lead to controlled flight into the ground, midair collision, or inappropriate control inputs resulting in aircraft stall (Patterson et al., 1997). The ubiquity of this problem has been well documented by mishap reports and surveys that indicate virtually all pilots' experience some form of SD during their careers. On an annual basis, accident statistics suggest each year SD results in the destruction of at least 20 DoD aircraft, the deaths of 25 flight crew, and asset losses of over 400 million dollars (Patterson, 2006).

Despite a lack of empirical accident investigation data, previous research has mainly emphasized anomalies within the vestibular system as the principal source of SD in flight. Consequently, most of the existing work on aviation SD has focused on nuances of the vestibular system under a variety of carefully controlled laboratory settings. Although, these efforts have generated direct applications toward treating clinical pathologies such as Meneire's disease or intravestibular trauma, directly linking these results to cockpit environments requires a considerable leap of faith. As an example, research evaluating the theoretical aviation vestibular illusion referred to as pilot "inversion" illusion revealed that producing this sensory misperception in the laboratory with a subject's eyes closed and head restricted, was relatively easy; however, in realistic cockpit environments with eyes open and head unrestricted, recreating this speculative SD problem was difficult if not impossible to accomplish (McCarthy 1994). Current accident statistics cast doubt on past efforts to mitigate SD through training and design efforts, since the SD mishap rate (25% -30%) has remained unchanged for decades and current trends suggest a possible increase in SD events (Patterson, 2012).

Although past explanations for spatial disorientation have concentrated on isolated vestibular illusions as primary causal factors, analysis of aircraft accidents and flight crew surveys suggest the most common forms of aviation SD are generated from conflicting visual cues that confound cognitive processing of pilot spatial strategies. This revised interpretation of causal factors has helped identify and classify common SD events, which has further led to specific SD categorization into interactive functional areas identified as vestibular, cognitive, and visual components *(Figure 1).*

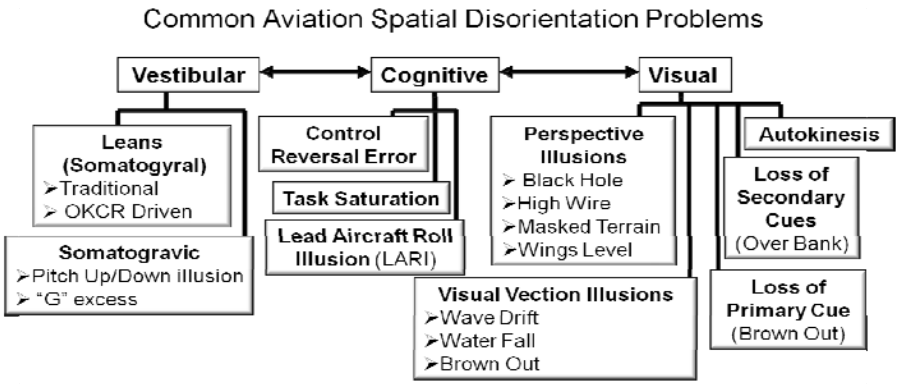


***Figure 1*:** Components of Aviation Spatial Disorientation.

Since current research suggests well defined pilot spatial strategies are essential for both prevention of SD and enhancement of cockpit designs, it is important to define this concept within the context of aviation nomenclature (Patterson et al., 1997). In general terms the word strategy refers to having a plan of action aimed at achieving a particular goal. When applied to spatial problems: a spatial strategy may be defined as a cognitive process that allows one to sense spatial patterns for the purpose of predicting and manipulating future spatial positions.

When describing in-flight spatial relationships, the term "sight picture" is often used to explain how visual cues interact within a pilot's field of view. To better characterize this descriptive technique, effective pilot "sight pictures" have been described as having specific retinal images defined as primary and secondary spatial cues. Patterson et al. (1997) have previously classified primary spatial cues as stabilized retinal images that remain centered on or near the pilot's central (foveal) field of view, while secondary spatial cues are defined as unstabilized peripheral images that will often appear in motion relative to the primary cue. During solo flight with visual meteorological conditions (VMC), distant steer points situated within the outside horizon view will typically serve as the pilot's primary spatial cue. Since perception of aircraft position, relative to this primary cue, is crucial for maintaining spatial orientation, peripheral views of the aircraft structures such as glare-shield, canopy bows, or wings play a vital role as secondary spatial cues that allow pilots to gage aircraft attitude relative to the horizon. Often times pilots will consciously, and sometime unconsciously, determine aircraft orientation by comparing retinal distances or angles between primary and secondary cue images (*Figure 2A*). The reasons why dynamics of these sight picture cues are critical toward formulation of an effective spatial strategy are: they provide immediate visual feedback indicating aircraft response to pilot control inputs. As an example, if a VMC solo pilot is using a real horizon as a primary spatial cue with the aircraft glareshield serving as a secondary spatial reference, a sight picture depicting increasing separation between these two cues would suggest the aircraft nose is dropping in response to forward control stick movements. Conversely, a decrease in separation between primary and secondary cues would provide a pilot with visual feedback indicating the aircraft's negative pitch angle was decreasing as the control stick was pulled backward. An important aspect of this spatial strategy is the concept that visual feedback from secondary spatial cues will move in the same direction as the stick control movements. This characteristic has been identified as a critical aspect of any human controlled system and is often referred to as the "principle of the moving part" (Roscoe, 1968).

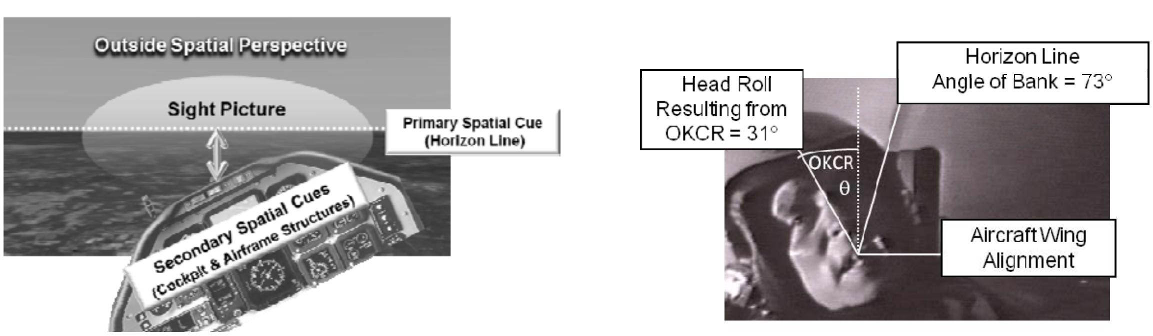


Figure 2: The image on the left (A) depicts primary and secondary spatial cues as they would appear to a pilot during a left bank maneuver and the image on the right (B) demonstrates a pilot exhibiting opto-kinetic cervical reflex (OKCR) head tilt toward the horizon during a high "G" 73 degrees left bank.

In contrast to conventional SD training and design doctrines, recent research has shown that sensory spatial reflexes routinely impact perception of spatial cues within a pilot's sight picture. Reflexes such as the opto-kinetic cervical reflex (OKCR) and opto-kinetic nystagmus torsion (OKN-T) have been shown to cause pilot head and eye rotation toward the horizon when looking "outside" the aircraft during VMC roll or pitch maneuvers (Patterson et al.,1997; Moore, 2008) (*Figure 2B*). Presumably, these intuitive sensory spatial reflexes improve spatial awareness by maintaining the horizon image as a retinaly stabilized primary spatial cue, against which peripherally viewed cockpit images (secondary spatial cues) will appear to move in synchrony with the control inputs.

In-flight Perspective Illusions

Perspective illusions occur in part because humans have a strong instinctive tendency to consider themselves level with distant horizontal references, regardless of their true altitude (Bresson, 2003). Also contributing to perspective illusions is the fact that anything we view as being slightly above our perceived horizon will appear as being magnified in size and exaggerated in its elevation. The strength of this illusion can be verified by viewing the common "moon illusion" (also known as the "Ebbinghaus illusion") which is a type of perspective illusion first identified in 350 BC (Roberts, 2005). This well-known spatial misperception occurs when a full moon located slightly above the horizon is perceived as being much larger than a full moon positioned directly overhead. In reality, the retinal images in both situations are the same; however, the human sensory-cognitive system has evolved in a manner that makes this illusion very convincing and difficult to overcome (*Figure 3A*). Although perspective illusions have been recognized and studied for centuries, the impact of these sensory misperceptions on aviation safety has only recently received close scientific scrutiny.

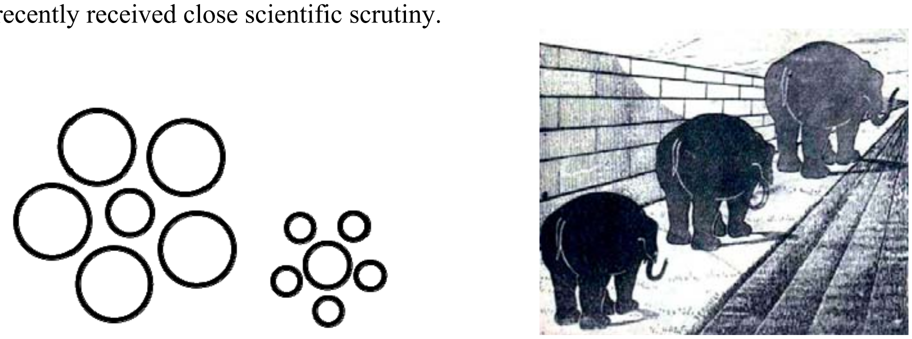


***Figure 3*:** (A) On the left is an example of the Ebbinghaus illusion, where the two central circles appear to have different diameters, but are the same size. (B) On the right is an illustration of the Ponzo (perspective) illusion; although the elephants appear different, they are all the same size. (http://www.moillusions.com/2008/09/ponzo-illusion-collection.html)

Among the confounding visual stimuli that can occur in flight, several visual perspective illusions have been identified as reoccurring mishap causal factors. Currently, these types of visual problems are defined as: "wings level", "masked terrain, "high wire", and "black hole" illusions. The etiology of cockpit perspective illusions revolves around the premise that pilots will on occasion, inadvertently apply a real horizon spatial strategy to situations where only fixed horizons or stationary ground references are available. Within low level flight environments, perspective illusions can readily occur when rising terrain or deteriorating visual conditions obscure the real horizon and, subsequently, force pilots to rely upon whatever fixed horizontal references become available within their forward fields of view. In these situations, distant shore or ridgelines are often used as horizontal references; even though fixed ground images of this type have spatial dynamics considerably different from those generated by real horizon images. The major cognitive-spatial differences between real and fixed horizons are: with level flight, the curvature of the Earth causes real horizon cues to move backward in space at a rate constant with the aircraft's forward movement. Because of this visual phenomenon, retinal image separation between the glareshield and horizon images will remain constant within a pilot's sight picture, as long as the aircraft attitude also remains constant. However, if aircraft attitude deviates in the roll or pitch axis, there will be a corresponding change in retinal distance between the glare shield and horizon images. With a real horizon, changes in the sight picture between the horizon primary and glareshield secondary cues provide a pilot with immediate visual feedback that indicates both the direction and the extent of pitch or roll variations. In contrast, when distant ground references are viewed from the cockpit, initial visual angles between the object and cockpit are usually quite small and somewhat similar to the visual angle of the real horizon. However, unlike a real horizon that keeps moving backward when an aircraft in level flight goes forward, distant ground references that are fixed in space will slowly descend within a pilot's sight picture as the distance between the aircraft and object decreases. The reason for this shift in visual perspective is: rather than remaining stable, the visual angle of ground objects changes at an increasing rate as the distance between the aircraft and object are reduced. When pilots have a real horizon available to serve as their primary spatial cue, changing visual angles of ground objects seldom create any spatial misinterpretations; however, when flying in conditions where outside visual references are degraded by weather or darkness, loss of a real horizon reference can create situations where pilots rely on fixed ground points as their primary spatial cue. Unfortunately, mishap statistics verify the following perspective illusions listed below continue to pose a significant threat to aviators during both good visibility day conditions and low visibility night landing approaches:

Wings Level Illusion – Accident investigators have reported that during low-level flights, *wings level illusion* can readily occur when rising terrain blocks the real horizon, thereby causing pilots to rely on whatever distant horizontal references are available within their forward fields of view (Patterson, 2007; Macknik, 2012). In this situation, distant shore or ridgelines will often appear as reliable horizon cues even though they are fixed or false horizons with azimuths that are usually several degrees below the true horizon. Since fixed horizon cues do not move backward in space as an aircraft in level flight transits forward, these types of horizontal references will gradually drop lower within the pilots visual field; subsequently generating a gradual visual angle decrease that will cause the primary cue (fixed horizon) and secondary cue (glareshield) retinal separation to contract within the pilot's sight picture. The spatial problem often encountered with this situation is: the gradually decreasing vertical distance between the fixed horizon cue and the glareshield will create the illusion that a level aircraft is nosing up and gaining altitude. In most cases, pilots flying through low level environments will compensate for this problem by frequently selecting new, more-distant fixed horizon references as they continue forward on the flight path; however, if an aviator in this situation attempts to use a fixed or false horizon as if it were a real horizon, the contracting sight picture may cause the pilot to react by slightly lowering the nose to keep the retinal distance between primary and secondary cues constant. This type of response will stabilize the sight picture (primary and secondary cue retinal distance) and generate the illusion that aircraft attitude is constant, even though the pilot may be unknowingly making slight increases in downward pitch to compensate for the changing visual angle between the glareshield and fixed ground reference. This *wings level illusion* is just one of several types of common perspective illusions encountered by pilots when they are forced to use a fixed, rather than real horizon.

Masked Terrain Illusion – When flying in low-level environments an additional form of the perspective illusion, referred to as *masked terrain illusion*, can readily occur when pilots are forced to rely upon fixed horizontal spatial cues instead of a real horizon. Similar to the *wings level illusion*, distant ridgelines often appear as reliable horizontal cues even though they are fixed or false horizons with azimuths that typically fall several degrees above the true horizon. Unfortunately, if there is a slightly lower ridgeline in front of a distant ridgeline being used as a horizontal reference, the front ridgeline will typically blend in visually (ie; become masked) with the more distant reference and may not become visually discernible until the aircraft reaches a point very close to the forward terrain. In the event a pilot is already experiencing *wings level illusion* while approaching a masked terrain area, ground impact may occur if a pilot fails to detect the forward obstruction within a recoverable reaction time. The terrain images in *Figure 4* replicate the primary cue perspective as it would have been viewed by a pilot involved in a recent *masked terrain illusion* aircraft accident. *Figure 5* illustrates the pilot's trajectory deviation attributed to keeping the sight picture constant while experiencing a wings level illusion.

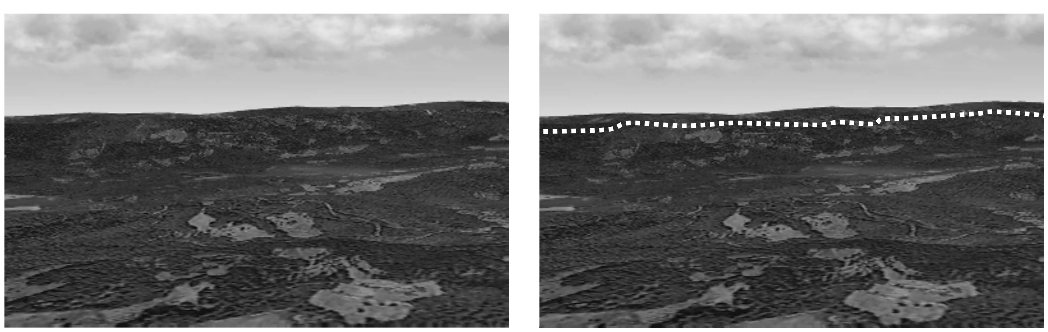


Figure 4: Left picture illustrates a front ridgeline visually blended with a more distant ridgeline. The right picture is the same image with a dotted white line added to indicate the location of the front (masked terrain) ridgeline.

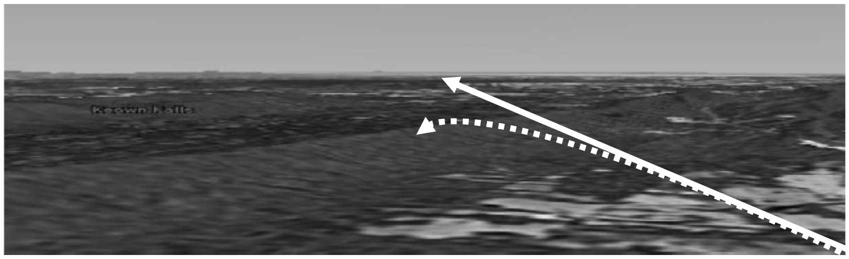


Figure 5: The solid white line represents the initial climb-out trajectory started by a pilot involved with a masked terrain illusion accident and the dotted line represents the mishap pilot's gradually reduced climb attributed to a combination of masked terrain and wings level illusions.

High Wire Illusion – Suspended wires are among the most significant threats faced by both military and civilian pilots (Patterson, 2010). Data from the National Transportation Safety Board indicates during a recent five year period (2005-2009) wire-strikes involving U.S. civilian pilots, damaged or destroyed 79 aircraft and caused 62 fatalities (Patterson, 2009). Within the U.S. Department of Defense, suspended wires also pose a major risk for military aviators that are often required to fly low altitude flights over unfamiliar terrain; U.S. Army reports indicate helicopter wire strikes average 3 to 4 per year and fixed wing aircraft within the Navy, Marine Corps, and US Air Force report wire strike averages of 2 to 3 per year (Nagaraj, 2008). Wire strikes can occur for reasons that range from inadequate planning to poor outside visibility; however, a specific perspective illusion identified as the *high wire illusion* has been attributed to several wire strike accidents involving both fixed wing and rotary wing aircraft. In several of these incidents, pilots gave similar description of this illusion by stating, "…the wires came out of nowhere, and looked like they were above us and rising. I only had a second to react by bunting the nose down to try to avoid them…". Characteristic of this illusion are reports that when wires are first spotted, mishap crews usually have very limited time to react (4-6 seconds) and in each documented case of *high wire illusion* the pilots believed they were at or below the wire altitude, when in actuality they were typically 100 ft. above the obstruction when it first came into view. If a suspended wire is positioned near or slightly below an aircraft's level flight path, its visual angle (as seen by the pilot) will be very small or near zero; subsequently the image of the suspended object will remain fixed or stable within the pilot's sight picture. In contrast, if the pilot is in a situation where it is necessary to use a fixed, rather than real horizon, the visual angle of the more distant and physically lower fixed horizontal reference will first slowly and then rapidly increase as the aircraft approaches (similar to wings level illusion describe previously). At a point dependent upon the aircraft's altitude and relative distances from the fixed horizon and suspended object, the suspended object (which will be centered on the fovea and therefore remain relatively stable in the sight picture) will appear to rise out of the visual ground clutter as the fixed horizon cue moves lower within the visual field. At this point, there is a high probability that a pilot will misperceive the suspended object as being above his aircraft and rising, even though the object is stationary, and may still be well below the aircrafts flight path. Unfortunately, in the limited amount of time available, pilots in this situation often instinctively react to this misperception by diving down to avoid what they perceive to be a high wire slightly above their flight path. Although, the *high wire illusion* is primarily generated from misperceptions of spatial cue movement within a pilot's sight picture, perspective lines such as those created by the sides of river banks or valley walls have been shown to further reinforce the *high wire illusion* via the Ponzo perspective illusion; which is known to amplify spatial misperceptions of object height and distance (Ganel, 2008 / *Figure 3B*).

Black Hole Illusion – Aviation spatial disorientation summaries often reference the visual phenomenon known as the black-hole illusion. The following excerpt provides what has become the conventional description of this common visual problem :*"A black-hole approach is one that is made on a dark night over water or unlighted terrain to a runway beyond which the horizon is indiscernible, the worst case being when only the runway lights are visible. Without peripheral visual cues to help him or her orient relative to the earth, the pilot tends to feel that the aircraft is stable and situated appropriately but that the runway itself moves about or remains malpositioned (is down sloping, for example). Such illusions make the black-hole approach difficult and dangerous and often results in a landing far short of the runway."* (Dehart, 1996). Recent pilot surveys and mishap statistics indicate excessively low black-hole landing approaches occur frequently among military pilots. Navy and Marine Corps flight crews report black-hole as the second most commonly encountered visual problem, surpassed only by misinterpretation of fixed horizon cues (Gallimore, 2003). Air Force surveys reinforce the severity of this cognitive threat by citing black-hole illusion as the leading visual problem for flight crews of multi-engine aircraft; and among all types of USAF pilots, the third most cited form of spatial disorientation (Matthews, 2002). On a dark night, with no visible "real" horizon present, pilots may unconsciously attempt to use their day/VMC spatial strategy by substituting the fixed ground cue of the runway as their primary horizontal reference. The problem with employing this spatial strategy at night is: unlike a real horizon which moves backward in space as the plane progresses forward, the viewing angle of the fixed runway image will slowly progress downward as the distance between the aircraft and the airfield continues to decrease (in a manner similar to the *wings level illusion*). Under these circumstances, the only way a pilot can maintain a constant "sight picture" is to slightly increase the aircraft's downward pitch angle (Patterson, 2011). Since this action is assumed to be an undetected reflexive response (rather than a well thought out process), it provides a credible hypothesis as to why pilots frequently enter into unintentional low and fast approaches during dark night landings.

Copyright statement

References

Bressan, P., Garlaschelli, L.,& Barracano M. (2003). Anti-Gravity Hills are Visual Illusions. *Psych Sci*, 14, 441-449.

Dehart, R. .L. (1996). *Fundamentals of Aerospace Medicine*, 2nd ed., Williams and Williams, p. 343.

Gallimore, J. J. & Ewart, R. (2003). A closed –loop system for effective spatial disorientation (SD) training. Department of Defense SBIR phase I technical report, N61339-03-C-0009/CLIN 0002AF, Sep 12.

Ganel, T., Tanzer, M., & Goodale, M. A. (2008). A double dissociation between action and perception in the context of visual illusions: opposite effects of real and illusory size. *Psych. Sci.* **19** (3): 221–5

Macknik, S, Marinez-Conde, S., & Gayles, E. (2012). Aviator's dilemma: pilots encounter illusions everywhere. *Scientific American/Mind*, June 4.

McCarthy, G. W. & Stott, J. R. (1994). In flight verification of the inversion illusion. *Aviat Space Environ Med.* Apr; 65(4):341-4.

Matthews, R. S., Previc, F., & Bunting, A. (2002). USAF spatial disorientation survey. Paper presented at the *Research and Technology Organization*/Human Factors and Med Symp on Spatial Disorientation in Military Vehicles: Causes, Consequences, and cures. April; La Coruna, Spain. HFM-085, RTO-mp-086.

Moore, S. T., MacDougall, H. G., Lesecux, S., Peyer, J. J., Wuyts, F., & Clark, J. B.(2008). Head-eye coordination during simulated orbiter landing. *Aviat Space Environ Med* ; 79: 888 – 98 .

Nagaraj, V. & Chopra, I., (2008). Safety study of wire strike devices installed on civil and military helicopters. *Department of Transportation* final report DOT/FAA/AR-08/25.

Patterson, F. R. (2012). Cockpit spatial strategies. *Proceedings:DoD HFE TAG*, (http://tg.hfes.org/sdtg/docs/sdtg_news_8-12.pdf)

Patterson, F. R. (2011). Breakthroughs in understanding aviation visual illusions. *Navy Medicine/ NAMRU-D Science Update* volume 1, issue 2.

Patterson, F. R. (2010). Aviator wire-strike avoidance training. *Research Information Bulletin* No. 10-3, Naval Aerospace Medical Research Laboratory.

Patterson, F. R., (2009). Unpublished NARML review of online data from (www.ntsb.gov/ntsb/query.asp.) Search terms included "wire strikes", "wires", and "guy wires".

Patterson, F. R. (2007) Human factors related to the EA-6B jet-Italian cable car tragedy: an expert witness account. *Aviat Space Environ Med*. ASMA (abstract-presentation) 78th Annual Scientific Meeting, May.

Patterson, F. R. (2006) Spatial Awareness Training Systems SBIR Phase II. *Naval Air Systems Command final report*. U.S. Government document, pp 1-58.

Patterson, F. R., Cacioppo, A. J., Gallimore, J. J., Hinman, G. E., & Nalepka, J. P. (1997). Aviation spatial orientation in relationship to head position and attitude interpretation. *Aviat Space Environ Med*, June: 68(6): p. 463-71.

Roberts, B., Harris, M. G., & Yates, T. A. (2005). The roles of inducer size and distance in the Ebbinghaus illusion (Titchener circles). *Perception.* 34 (7): 847–56.

Roscoe, S. (1968). Airborne displays for flight and navigation, *Human Factors*, 10, 321-32.

DESIGNING FOR JOINT HUMAN-AUTOMATION COGNITION THROUGH A SHARED REPRESENTATION OF 4D TRAJECTORY MANAGEMENT

Rolf Klomp, Clark Borst, M.M. (René) van Paassen, Max Mulder
Department of Aerospace Engineering, Control & Simulation Division,
Delft University of Technology, Delft, The Netherlands
Dennis Nieuwenhuisen, Anneloes Maij
National Aerospace Laboratory (NLR), Amsterdam, The Netherlands
Martijn Mooij, Annemiek van Drunen,
Thales Research and Technology, Delft, The Netherlands

The current evolution of the ATM system, led by the SESAR programme in Europe and the NextGen programme in the US, is foreseen to bring a paradigm shift to the work domain of the air traffic controller. A focal point is the introduction of the 4D (space and time) trajectory as a means for strategic management rather than the current –hands on– method of control. In both programmes a central role is foreseen for the human operator, aided by higher levels of automation and advanced decision support tools. However, many other complex socio-technical domains have shown that the transition to higher levels of automation often introduces new problems, problems that are harder to resolve than the ones intended to solve in the first place. This paper presents one approach to the design of a shared representation for 4D trajectory management. The ultimate goal is to design a shared representation which forms the basis for both the design of the human-machine interfaces and the rationale that guides the automation. It is expected that such a shared representation will greatly benefit the joint cognition of humans and automated agents in ATM and will mitigate breakdowns in coordination by design. A preliminary version of a joint cognitive representation for 4D trajectory management has been developed and is introduced in this paper. Future work will focus on the further development and refinement of shared representations by means of human-in-the-loop experiments.

THE current evolution of the air traffic management (ATM)-system is expected to result in a situation where high-precision four-dimensional (4D, i.e., space and time) trajectories for aircraft, stored in automated support tools, will form the basis for the work of the human controller. A pull is provided by the increasing demand which is placed on the Air Traffic Management (ATM)-system (e.g., workload, capacity, efficiency, etc.). Conversely, technological advances on the air- and ground side of the ATM-system (e.g., advanced flight management systems, high precision trajectory prediction algorithms, domain-wide communication systems) provide a push facilitating a new form of Air Traffic Control (ATC).

A fundamental shift in the work domain of the human Air Traffic Controller (ATCo) is the introduction of *time* as an explicit control variable for defining *aircraft (4D-)trajectories*, rather than the current tactical –hands on– form of control. Whereas the current form of ATC mainly relies on the skill and experience of the human controller and is often performed with little help from automated tools, a shift towards 4D-based ATM will no longer be possible without the aid of advanced automated support- and decision-making tools (Parasuraman, Sheridan, & Wickens, 2000).

Although considerable research has been devoted to exploring this future approach of ATC with 4D trajectory support, a definite breakdown of the distribution of roles and coordination between the human operator and automation and the 'central role' of the human operator is not yet well defined.

From other complex socio-technical domains it has been shown that the transition towards higher levels of automation often introduces new problems, problems which are often harder to resolve than those intended to be solved in the first place (Bainbridge, 1983). Increasing the level of automation is –in itself– not good or bad. However, with increased automation a more extensive form of coordination between humans and automated systems will be required (Christoffersen & Woods, 2002).

Breakdowns in this coordination may result in humans having difficulty to getting the automation to do what they want, and conversely, a poor understanding of how the automation works (Rasmussen, 1986). To facilitate the coordination between human- and automated agents, it is imperative to create new forms of automation and to make the automated systems 'team players' *by design*.

This paper outlines an approach for designing shared human-automation cognition for ATM, based upon 4D trajectory management. The work presented is conducted in the context of SESAR WP-E project 'C-SHARE'. Furthermore, a first prototype of a Joint Cognitive System (JCS) (Hollnagel & Woods, 2005) will be introduced, illustrating how this design approach can be used as a basis for ecological Human Machine Interface (HMI) design.

Designing for Shared Cognition

The introduction of higher levels of automation is essential in order to facilitate the shift towards a new form of 4D air traffic control. On one hand, advanced automated support tools are necessary to aid the human operator to cope with the increased complexity of 4D trajectory management. And on the other, as a systems designer there are many valid reasons to take advantage of modern computational technologies (e.g., data fusion and information processing, advanced algorithms). For example, to optimize the global use of airspace, reduce traffic complexity and minimize the cost for individual airspace users.

From other complex socio-technical domains however it is clear that the introduction of a higher level of automation in itself does not guarantee an improvement in overall system performance. There is an abundance of empirical, operational and theoretical evidence that breakdowns in human-automation coordination can introduce severe human and systems problems contributing to incidents and accidents, including transient workload peaks, 'out-of-the-loop' situation awareness and vigilance problems, overreliance (complacency), and skill degradation. Therefore, it is imperative to mitigate these breakdowns in coordination by design and make the automated systems 'team players'.

When looking at effective human-human interaction in productive team thinking and problem solving, its foundation for success is a shared understanding –or a 'common ground'– of the task to be achieved and the paths by which this task can be achieved. Similarly, when looking at human-automation coordination, a common ground can be found in the properties, functions and constraints active on the work domain (or problem space) to which all actors must abide. It is to be expected that when these elements are somehow made visible to the human operator(s) and conversely act as a basis to guide the rationale of the automated agents, breakdowns in coordination can be mitigated and productive collaboration and team thinking can be achieved.

One systematic approach to identify this common ground for effective human-automation coordination is given by the framework of Cognitive Systems Engineering (CSE) (Hollnagel & Woods, 1999). Contrary to the user- and automation centred design approaches which take the needs, wants and limits of respectively the human operator(s) and automation as a starting point for design, CSE is based upon the global context in which the work takes place (i.e., the work domain), irrespective of any definite systems design or pre-determined task allocation.

As a first step in CSE, a functional breakdown is made of the work domain, identifying all relevant elements and functions on various levels of abstraction. As a subsequent step, and based upon the reasoning that knowledge of the entire system cannot be solely be built up from knowledge of the individual parts, the underlying relationships between the elements which define the global context are sketched using means-end links; basically asking the question of "how does it work?" and "why is it here?" for each element (Rasmussen, 1986).

When considering the work domain for air traffic control, various phases for the refinement of 4D trajectories are foreseen; from long term seasonal planning to the in-flight revision of trajectories during the tactical monitoring phase. It is foreseen that in each phase a unique form of coordination will exist between the human operator, their displays and support tools, and automated agents (Van Paassen et al., 2011).

For the scope of this research, focus has been put on designing a framework for shared cognition in the *tactical monitoring phase*, the in-flight management of 4D trajectories by ATC, as it provides the most challenging environment for human-automation coordination (e.g., time-critical, safety-critical, high dynamic complexity, and 'open' (Rasmussen & Pejtersen, 1990) work domain). Contrary to any prior planning phases which are deterministic in nature, the main task of the human operator in the tactical phase will be to identify and effectively cope with *any* unforeseen events.

Following the first step of CSE, an initial Work Domain Analysis (WDA) has been performed for the scope of the research by the construction of an Abstraction Hierarchy (AH) and is shown in *Figure 1*. Furthermore, a breakdown is given of the abstract functions in the work domain.

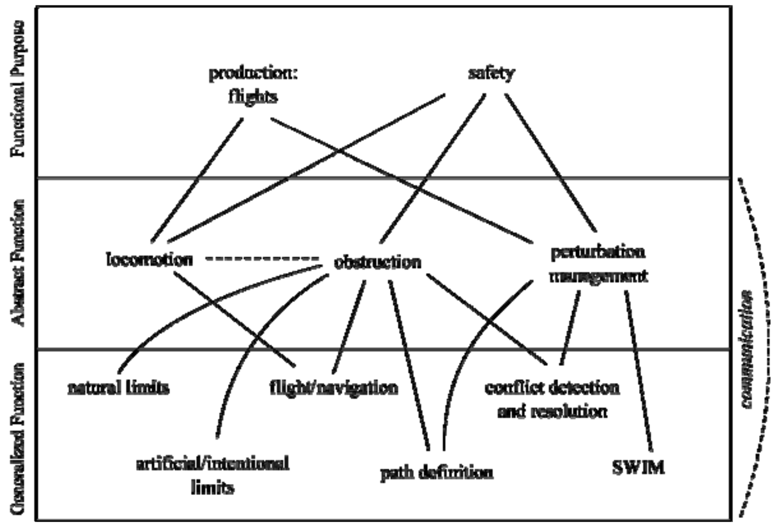


Figure 1. The top three levels of the Abstraction Hierarchy for the tactical monitoring phase.

Locomotion is realized within the constraints following from individual flights and their respective navigation within the environment. These constraints can be imposed by *internal* factors such as the aircraft performance envelope and the availability and fidelity of navigation systems, but also by *external* constraints such as airspace user preference and airspace regulations. The absolute locomotion of the moving agents can be captured in their resulting (4D-) path definition. The relative locomotion is then, in turn, realized by the dynamics of travel of all agents within the system.

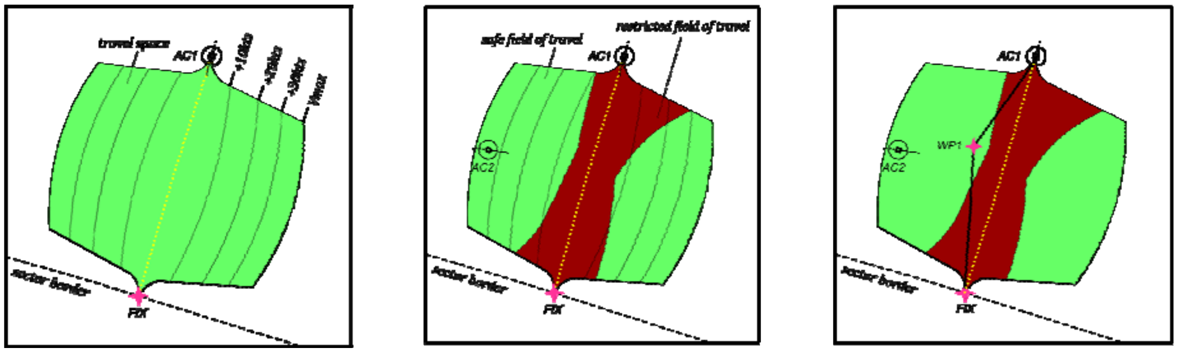
Obstruction is realized by both static and dynamic constraints which limit the system from performing at a theoretical optimum. Such obstructions can be in the form natural limits (terrain, weather, …) and artificial constraints (separation minima, airspace structure, …). Furthermore, the relative locomotion of all moving agents and the current network status also impose obstructions on the operations.

Perturbation Management is realized by the awareness and integration of the intended-, current and projected state of the work domain. Here, the main source of the information (path definitions, Network Operations Plan (NOP) status, meteorological information, …) for both the human and automated agent is foreseen to follow from a system wide information system. Furthermore, conflict detection algorithms are foreseen to provide more detailed information about safety critical perturbations.

Although the AH highlights the underlying functions which govern the work domain, it does not provide a final recipe for how shared cognition can be obtained through a specific human-machine automation design. Determining which form of representation including its interaction with both the human user and automated agents is suitable is still a creative step and depends on the (sub-) task for which it is designed. However, the functional breakdown in the AH provides guidance in determining which functions, constraints and relationships should somehow be made visible in a shared representation.

Travel Space Representation

As a starting point for the design of a joint cognitive system, a prototype of a constraint-based shared representation has been designed for the task of the in-flight manipulation and revision of 4D trajectory by ATC. According to definition a 4D trajectory consists of a set of consecutive segments linking 4D points (waypoints), at which the indicated times are estimates in the form of target times or times subject to constraints (Eurocontrol, 2007). The manipulation and placement (position and timing) of such waypoints is taken as the task to be shared between the human users and automated agents. Re-planning of waypoints is necessary in case one or more inherent (other traffic, terrain, weather, …) or intentional (restricted airspace, procedures, …) constraints active on the aircraft trajectory, cannot be satisfied due to any number of unforeseen events.



(a) Travel space representation for a single aircraft

(b) Traffic causes restricted field of travel (loss of separation)

(c) Placement of intermediate waypoint to ensure separation

Figure 2. Conflict resolution by trajectory re-planning with the travel space representation

For the design of a shared representation the Ecological Interface Design (EID-) (Vicente & Rasmussen, 1992) approach has been adopted. EID closely follows the line of CSE, and argues that for effective human-machine interface design the constraints and underlying relationships governing the work domain (or ecology) should be somehow made visible to the human operator. It is hypothesized that, by visualizing the task-relevant functional constraints that arise from the work-domain, the *same* constraints that guide automated actions, humans will get a deeper understanding of why automation proposes a particular action. This may benefit the operator's trust and acceptance of the automation, and facilitate the transition back and forth from higher levels of automation.

Safe field of travel

When considering the task of in-flight re-planning of a 4D trajectory by ATC the relevant functions and constraints which govern the work domain can be derived from the AH. The overall goal of the aircraft is to execute its subsequent trajectory segments and pass all waypoints within the timing constraints in agreement with ATC. Now consider that either the aircrew or air traffic controller intends to introduce an intermediate waypoint into a trajectory segment. Any arbitrary (4D-)placement of that waypoint will lead to a new definition of the trajectory. The feasibility of such a trajectory can be tested against the relevant constraints which govern the work domain (e.g., adherence to *locomotion*, *obstruction* and *perturbation management*). Then, the subset of all trajectories which adhere to these constraints are feasible solutions and, by definition, form a so called 'safe field of travel'. By means of one-to-one mapping, a correspondence-driven (Vicente, 1990) translation of this safe field of travel can be made on the air traffic controllers' plan view display, indicating the real-world spatial locations of feasible waypoint placements and their timing implications.

Representation breakdown

In *Figure 2(a)* the basic composition of the travel space representation is shown. Aircraft *AC1* is flying along a pre-agreed 4D trajectory towards a certain metering fix (point *FIX*) at the sector border. The Controlled Time Over (CTO) at the fix is taken as a hard constraint (i.e., it must be met). When considering constraints which follow from the aircraft performance envelope (in combination with the time constraint at the fix), an area can be bounded in which intermediate waypoint placement is feasible. The aircraft turn characteristics determine the rounded shape of the travel space close to the current aircraft position and the metering fix. Furthermore, any intermediate waypoint that does not lie directly on the current trajectory segment implies an increase in track length, and thus an increase in required ground speed. The outer edges of the travel space are therefore bounded by the maximum achievable speed within the aircraft performance envelope.

In *Figure 2(b)* other traffic has been introduced in the form of a single second aircraft (*AC2*). When taking the separation constraints for *both* aircraft into account, an area within the travel space for *AC1* becomes restricted (i.e., an intermediate waypoint in that area will result in a 4D trajectory which is in conflict with the other aircraft at a certain point in time). This area is indicated in the figure as the *restricted field of travel*. The restrictive area visualizes the locations where an intermediate waypoint would not lead to the resolution of the conflict. Note that these same constraints on waypoint placement *also* hold for an automated agent. It is essential to understand that these constraints arise from the work domain, *independent* of who will act on the task of resolving the conflict, the human, the automation, or both.

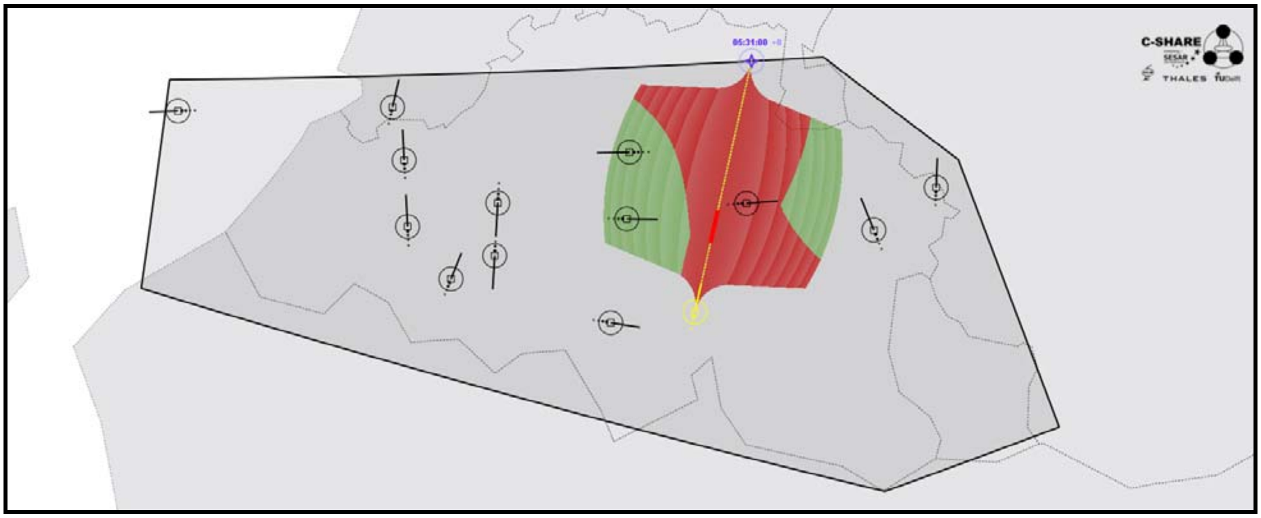


Figure 3. Interactive software based implementation of the travel space representation

Figure 2(c) shows how the travel space representation can be used by the human- or automated controller to select an appropriate position for an intermediate waypoint in a conflict situation. By placing the waypoint (*WP1*) inside the safe field of travel within the travel space, the constraints following from aircraft performance, separation, and timing are all met. Note that here, the timing of the introduced waypoint is set such that it corresponds with the constraints visualized by the representation (e.g., constant speed along both segments and fixed timing at the final waypoint).

This visualization of the work domain constraints and their relationships allows a human controller to reason about, and directly act upon the airspace environment. To emphasize once again, note that this same representation can be used to guide the rationale of an automated agent or, equivalently, a team of human operators and automated agents to achieve productive collaboration and team thinking. For example, an automated agent could propose a resolution and map this resolution within the safe field of travel. By carefully observing the machine's advisory, the human agent could either 'accept' or 'veto' the advisory warranted by the demands of the situation at hand.

In other words, users are not only able to see the intentions of the automated agents, but they are also able to re-direct machine activities easily in occasions where they see a need to intervene. By visualizing the task-relevant functional constraints that arise from the work-domain, the *same* constraints that limit automated actions, it is hypothesized that humans will get a deeper understanding of why automation proposes a particular solution. This may benefit the operator's trust and acceptance of the automation, and facilitate the transition back and from higher levels of automation.

Discussion

Perhaps the largest change for an ATCo in future ATM operations will be to step away from the current hands-on tactical control of aircraft to an operation in which traffic is planned in detail beforehand. For individual flights, it has proven possible to implement, monitor and manipulate 4D trajectories, usually in the context of all other aircraft being controlled traditionally. The case when all aircraft are to be controlled based on their 4D trajectory means a tremendous step, and a real-time visualization of how all trajectories will evolve in time is a big challenge for display designers. Whereas the dimensionality of the control problem explodes, the visualization and display techniques remain limited by, among others, clutter issues, and physical constraints such as screen size and resolution.

The prototype of the travel space representation has shown to be a good starting point; this framework for 'joint cognition' can act as a basis for designing both the automation support and the human-machine interfaces, in the air and on the ground, from one and the same perspective. A qualitative evaluation with an initial interactive software based implementation of the travel space representation (*Figure 3*) has shown that this type of representation is indeed suitable as a common ground for reasoning between the human operator and automated agents in resolving *local* perturbations (e.g., re-plan the trajectory of individual aircraft). However, one can argue that in case of larger scale perturbations (i.e., regional or network-disruptive) the complexity and temporal load of focussing on the constraints of individual aircraft could exceed the cognitive limits of the human controller. In that case perhaps, a common ground should be found in the higher level properties of the airspace such as intrinsic complexity, robustness to perturbations, flexibility and the flow of traffic. It is hypothesized that the controller could then –by means of a more heuristic approach– structure the global use of the airspace and traffic, and zoom in for detailed refinement where necessary. This is foreseen to result in a situation where humans will have a deeper understanding of the actions and reasoning governing the automated agents, and will facilitate the transition back and from higher levels of automation.

During the further development and testing of prototypes, it is likely that the Work Domain Analysis will need to be augmented and/or partially revised. A number of human-in-the-loop experiments are foreseen that will show to be crucial in converging the design and analysis iterations to a representation of 4DT management and that can indeed be used for both automation and human-machine interface design.

Acknowledgements

The authors acknowledge the inspiration from EUROCONTROL and the SESAR Joint Undertaking. The work was co-financed by EUROCONTROL on behalf of the SESAR Joint Undertaking in the context of SESAR Work Package E (project C-SHARE: Joint ATM Cognition through Shared Representations). This work reflects only the authors' views and EUROCONTROL is not liable for any use that may be made of the information contained herein.

References

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), pp. 775–779.

Christoffersen, K., & Woods, D. D. (2002). How to Make Automated Systems Team Players. (E. Salas, Ed.) *Advances in Human Performance and Cognitive Engineering Research*, *2*(2002), pp. 1–12.

Eurocontrol. (2007). *SESAR Concept of Operations* (p. 214). Brussels, Belgium. doi:DLM-0612-001-02-00

Hollnagel, E., & Woods, D. D. (1999). Cognitive systems engineering: new wine in new bottles. (W. Karwowski, Ed.) *International journal of humancomputer studies*, *51*(2), pp. 339–356.

Hollnagel, E., & Woods, D. D. (2005). *Joint Cognitive Systems*. (E. Hollnagel & D. D. Woods, Eds.) *Joint Cognitive Systems Foundations of Cognitive Systems Engineering*, pp. 113–133, Taylor & Francis Group, LLC.

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), pp. 286–297.

Rasmussen, J. (1986). *Information processing and human-machine interaction*. (A. P. Sage, Ed.) *An approach to cognitive engineering*,Vol. 12, p. 228, Elsevier Science Publishers.

Rasmussen, J., & Pejtersen, A. M. (1990). Taxonomy for Cognitive Work Analysis. *Analysis*, (September), p. 155.

Van Paassen, M. M., Borst, C., Mulder, M., Klomp, R., Van Leeuwen, P., & Mooij, M. (2011). Designing for Shared Cognition in Air Traffic Management. *SESAR First Innovation Days*, pp. 1–5.

Vicente, K. J., & Rasmussen, J. (1992). Ecological Interface Design: Theoretical Foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, *22*(4), pp. 589–606.

Vicente, K. J. (1990). Coherence- and correspondence-driven work domains: implications for systems design. *Behaviour & Information Technology*, *9*(6), pp. 493–502.

ARRIVAL MANAGEMENT DECISIONS BY VISUALISING UNCERTAINTY

M. Tielrooij, C. Borst, M.M. van Paassen, M. Mulder
Faculty of Aerospace Engineering, Delft University of Technology
Kluyverweg 1, 2629HS, Delft, The Netherlands

To balance the flow of inbound aircraft and the capacity at airports, more and more Air Navigation Service Providers use Arrival Management (AMAN) systems. These provide decision support to sequence managers in planning inbound flights to optimize capacity, flight efficiency, and predictability. All AMANs are based on predictions of an aircraft's arrival time. Due to various disturbances the error of these predictions grows larger with the prediction horizon. Air Traffic Controllers will therefore not be able to effectively use the support at a certain horizon due to the lack of confidence in the provided information. This paper proposes and tests an enhancement based on the probability density function of the expected arrival time, allowing controllers to include uncertainty in the decision making process.

Arrival Managers (AMANs) aim to balance the inbound flow of aircraft with the available capacity at the airport. When aircraft are predicted to arrive too close after each other, AMAN provides support in deciding how to influence the 4D trajectory of the aircraft involved. Assuming aircraft fly an optimal trajectory from the operator's perspective, disturbance from this trajectory should be kept to a minimum. For example, a smaller speed increase over a longer flight time is more fuel efficient than a larger speed increase of over a shorter time, while achieving the same difference in time. To improve performance and to support future trajectory-based operations, the planning horizon of AMAN is envisaged to be increased from the current typical 100 NM to 200-500 NM (Barff et al., 2012; Bronsvoort et al., 2011). This planning horizon is currently limited by the ability to get information on the aircraft at a longer horizon, the ability to influence the aircraft further from their destination, and the reliability of the predicted arrival times.

System Wide Information Management will enable continuous sharing of all relevant information concerning a flight between all involved actors (SESAR JU, 2011; JPDO, 2008). Through SWIM, different Air Navigation Service Providers (ANSPs) will also be able to share their requirements on a trajectory (such as an arrival time planned by AMAN) as well as their capabilities to provide for such requirements. While this resolves the first two limitations on the planning horizon, future prediction uncertainty is expected to reduce but unlikely to disappear altogether. As disturbances may influence the trajectory over a longer time and new sources of error may be introduced, a longer horizon will increase the uncertainty in arrival time (Mondoloni, Paglione, & Green, 2002; Mueller, Sorensen, & Couluris, 2002; Hunter 2004).The actual uncertainty may vary due to for example weather, actual traffic, or aircraft navigation capability. However, the uncertainty is transparent to the operator; the effective horizon will be based on general experience and not on the actual accuracy at a given time.

Research has demonstrated the ability to calculate the uncertainty in a particular trajectory (Whysall, 1998; Mueller et al., 2002; Schaefer, Gizdavu, & Nicholls, 2004). It is hoped that, by providing this information to the human operator, a higher benefit from AMAN may be achieved in situations with low uncertainty through better decision making.

This paper consists of three parts. The first two sections describe the development of the visualisation of the information on a common concept for AMAN using the Ecological Interface Design (EID) framework. Secondly, the interaction with such an interface is discussed leading to further definition of the visualisation. The last segment describes the setup, and execution of initial experiments to test the visualisation concept.

Approach

Most current AMAN display interfaces are based on a moving timeline on which the expected or planned arrival times are shown (Hasevoets & Conroy, 2010). This allows a 2D representation of the 4D spacing problem as relevant to the planner (who is not separating the traffic in 3D but rather adjusting the flow to allow for easier separation). However, this display provides limited information on the available capacity, the required spacing, the limits on the accuracy of the arrival time, or an aircraft's capability to meet that time. The lack of such information results in either a need for extensive knowledge and experience of the planner, or is not accounted for leading to lower overall performance.

The EID framework (Vicente & Rasmussen, 1992) helps in developing of the display that presents the content and structure of the working environment using the Abstraction Hierarchy (AH). Subsequently, the form of presentation is developed using the skills, rules, and knowledge taxonomy (Rasmussen, 1983). A display has been designed that explicitly visualizes the uncertainties in arrival times, and how these uncertainties would affect sequencing performance and controller strategies/control actions. Thereby a constraint-based approach, inspired by the EID paradigm, has been adopted.

Visualisation

Current timeline presentations show the Estimated Time of Arrival (ETA), or the planned time of arrival. None of the current operational systems, and very few of research systems, shows the required spacing between two aircraft. No explanation for the lack of this information could be found in literature. This parameter is however the key factor in safety (i.e., minimal amount of spacing) and capacity (i.e., available room for spacing). In the concept display, the required spacing is shown as a blocks, see Figure 1(a). The block indicates the time that the aircraft occupies the available landing capacity. Its surface then represents demand (expressed in seconds). A single aircraft would use a capacity of 1 (i.e., 1 runway) for that amount of time. This provides the separation requirement and runway occupancy but, as Figure 1(a) shows, provides poor indication of a lack of separation when these blocks overlap.

If a single block represents the use of a landing slot (in time), blocks of different aircraft may be added up to indicate instantaneous demand at each moment in time. Any demand higher than the number of available landing slots represents a shortage of capacity; see Figure 1(b). In this form, the equality (during the planning phase) of a predicted loss of spacing and a shortage of capacity is directly evident. Both of these issues require action, and any solution to one problem also resolves the other.

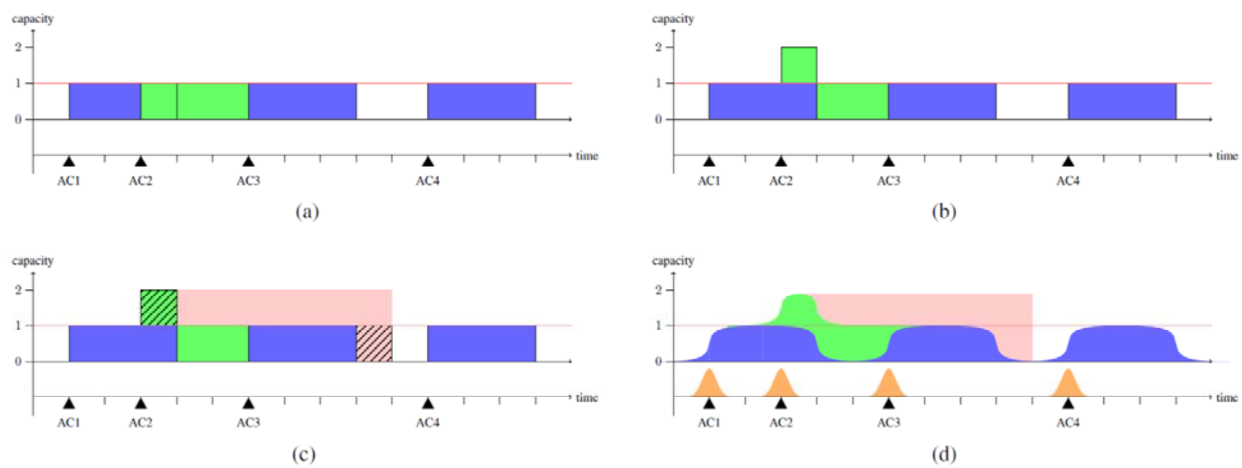


Figure 1: Development stages of the enhanced timeline. The blue area shows the amount of capacity used by the aircraft, the orange shapes show the arrival time PDFs, and the pink area shows the relation between an excess in demand and the earliest time that that excess demand can be resolved.

Arrival Time Uncertainty

Uncertainty information is added by providing the Probability Density Function (PDF) of arrival over time on the timeline; see Figure 1(d). The width of this graph indicates the time at which the aircraft may arrive, the shape indicates the most likely arrival time (the highest point) as well as the likelihood that the aircraft may deviate from that Development stages of the enhanced timeline. The blue area shows the amount of capacity used by the aircraft, the orange shapes show the arrival time PDFs, and the pink area shows the relation between an excess in demand and the earliest time that that excess demand can be resolved.

When the ETA is provided as a PDF, the nature of the occupancy blocks should change as well: When an aircraft is predicted to have a given probability on a given time, it has an equal probability of occupying the separation interval from that time. This results in a Cumulative Density Function (CDF) limited by the separation time (visible in Figure 1(d) as the blue shapes). Effectively, the occupancy block now becomes the expectation value for the occupancy for each aircraft.

Resource Occupancy

As the runway is occupied for a certain amount of time after the aircraft arrives, the instantaneous expectation for runway occupancy due to the expected arrival at this time can be expressed as:

$$O_{P_i(t)}\left[t, t+s\right] = P_i\left(t\right)$$

In which s is the applicable spacing interval. The expectation for the occupancy at a given time then becomes the integral or ---the CDF--- of the arrival time probability for the spacing interval before it:

$$O_i\left(t\right) = \int_{u=t-s}^{t} P_i\left(u\right) du$$

The CDFs for all aircraft can again be added up to calculate the total expected occupancy:

$$O\left(t\right) = \sum_{i=1}^{n} \int_{u=t-s}^{t} P_i\left(u\right) du$$

The summed CDFs provide the expectation value for runway occupancy for all aircraft. When the CDF is equal to the capacity, the runway will be occupied, regardless of which aircraft will be occupying it. Even in situations with very high uncertainty, this indicator can provide support in balancing capacity to demand without yet knowing the exact landing sequence of the aircraft.

Interaction

When the aircraft are indicated at their respective ETAs, the display only provides the current predicted arrival schedule. To support the controller in developing a suitable planning the controller can test potential modifications to the schedule directly on the display. These probes are implemented as a Direct Manipulation Interface (DMI) (Hutchins, Hollan, & Norman, 1985).

The operator is able to directly modify the arrival time of aircraft, and by doing so, see the potential effect of the change on the situation. The system provides real-time update of the arrival time PDF and CDF, therefore showing the complete expected result of the action.

The direct manipulation style of human-computer interaction is particularly useful for probing different solutions as it immediately shows whether a solution is furthering the goals of the user. By highlighting the CDF of the aircraft being probed (in Figure 1(d), the contribution of the selected aircraft is shown as the green area), the user can also explore the contribution of the aircraft to the total capacity problem.

Planning Limits

Not all possible changes in arrival time are available: aircraft have a limited maximum speed, which limits the amount of time before the ETA that aircraft can arrive. Similarly, aircraft have a limited endurance, which limits the maximum delay. Furthermore, either deviation will at some point no longer be efficient for the aircraft operator, and therefore undesirable. In combination with DMI, these limits can easily be visualised as an interval in which the aircraft can be planned.

Occasionally it may be that the required delay for the last aircraft in a series of too closely packed aircraft is not feasible. In such a case either the first aircraft in the sequence have to arrive earlier or extra capacity (runways) needs to be made available. If aircraft can only be planned one at a time, the required delay on the last aircraft can only be determined once all earlier aircraft have been delayed sufficiently.

To assist the planned in deciding on whether to advance aircraft, delay aircraft, or add capacity, the total amount of required delay is shown. Figure 1(c) shows the graphical relation between such an excess in demand, and the earliest time at which it is resolved; as soon as an equal area above the capacity limit is provided as unused area below that limit, the problem is resolved. The planner can now directly see the amount of delay required on the last aircraft of a sequence. Figure 1(c) shows AC3 to be the last aircraft involved. If this delay is too large, the solution has to be found elsewhere. Similarly, any aircraft beyond the area is unaffected by this problem. As the determination of this time to resolution is solely based areas in the graph, Figure 1(d) demonstrates that the same technique can also be used in the uncertainty display.

Experiment

Note that the display described above is has no attributes specific to aviation. The display supports solving a planning problem in which certain actors will use a certain limited resource for a specific amount of time, at a specific time in the future. Therefore, the display might be applicable to other logistic planning problems such as shipping or railways for example.

The lack of specific context also allows testing the display with untrained human subjects rather than operational experts. This display was tested with seven students aged 22 to 28, who, while all having a background in aerospace engineering, have no operational background in Air Traffic Control.

The experiment's objective was to determine the effect of the addition of uncertainty information and the resolution information on the ability to efficiently plan inbound traffic. In the experiment, subjects were provided with four different displays: the block-type display without uncertainty, PDF-based display, and both displays with the indicator on resolution time (Figure 2).

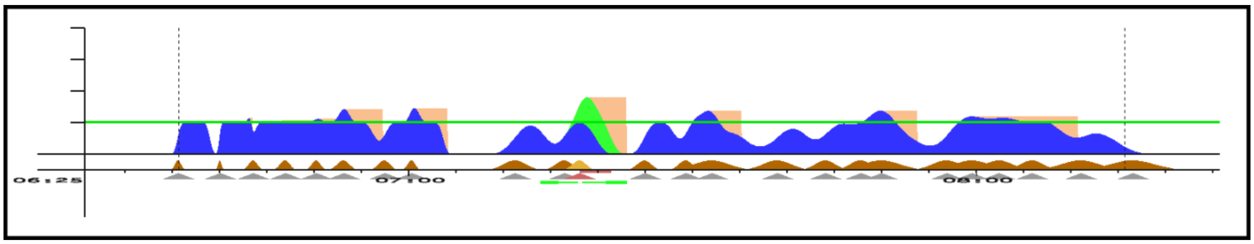


Figure 2: The experiment display in the full configuration. For clarity, the background colour has changed from black to white. The green lines below the aircraft symbol show the control space available for that aircraft.

Subjects were tasked with spacing traffic on a horizon of 2 hours with a minimal separation of 200 seconds between each aircraft. Subjects were asked to monitor and ensure sufficient separation first of all. Secondly, their task was to give instructions as early as possible, and to minimise the number of instructions for each aircraft. Since this would be a very low workload task with limited measuring possibility, traffic was sped up 30 times allowing for 3 hours of traffic to be simulated in runs of 6 minutes.

The scenarios were set up to have an running average spacing over 5 aircraft of 270 seconds at arrival to provide adequate solution space with sufficient aircraft landing to determine performance. Subsequently, a prediction error was superimposed on the actual arrival time. The prediction error was based on a normal distribution with a variable initial standard deviation that reduced to a fixed, small final standard deviation. Aircraft were modelled to fly constant, identical speeds to the runway with an ability to accelerate or decelerate an equal amount.

Each subject completed 8 training runs, followed by 16 measurement runs in which the combination of display and scenario was Latin-squared to eliminate training or fatigue effects. To determine the subjective workload of the planner, an Instantaneous Self-Assessment (ISA) probe appeared on the side of the screen every 30 seconds (Tattersall & Foord, 1996). The simulation system further recorded all changes to the planning and the resulting landing schedule.

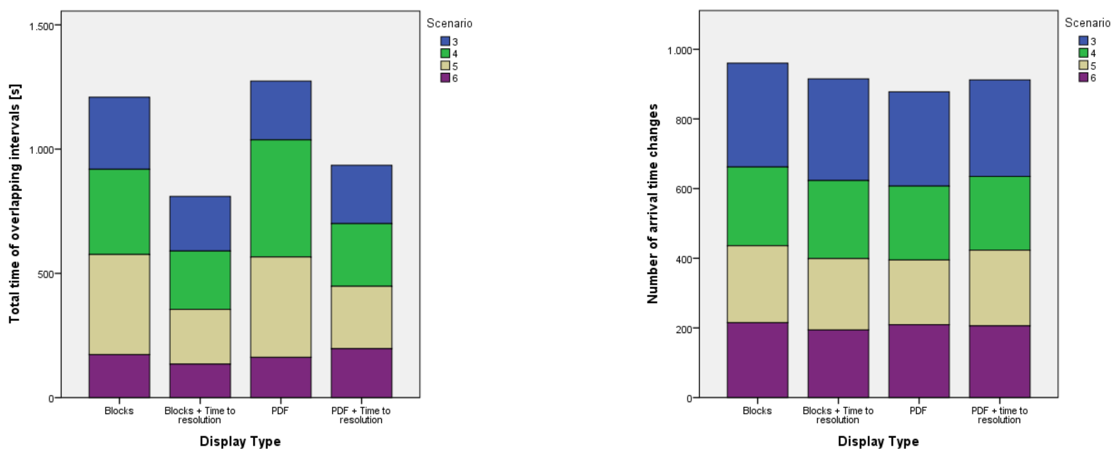
It was hypothesized that the new display would allow for more gradual planning in which the spacing buffer is adjusted to suit the uncertainty of the aircraft involved. This in turn should lead to less occurrences of predicted overlap resulting in fewer corrective actions on spacing and less spacing conflicts at landing.

Results

Initial analysis of the workload rating showed a clear correlation between duration of the experiment and perceived workload, even at the later experimental runs. This coincided with comments from all subjects that they only became comfortable with visualisation of uncertainty at later stages of the experiment. No further trends or effects could be found on the presentation of uncertainty, suggesting that the training stage was too short to effectively use the new visualisation.

Comments from the experiment subjects indicated that the uncertainty display was considered complex. In particular understanding the contribution of the each aircraft to the total CDF was considered unpredictable. This may be due to the morphing shape as uncertainty decreased. In general subjects preferred the blocks as they provided a more direct indication of the amount of buffer between two slots. Subjects did indicate that the size of the uncertainty helped them in estimating the required amount of buffer between two aircraft.

In both displays with the resolution time indication, the number of remaining spacing conflicts, and the total time of overlap was considerably lower than in their baseline counterparts, as shown in Figure 3(a). Figure 2 shows that the indicator could help in highlighting small conflicts by introducing a new colour on the screen. To eliminate the possibility that this indicator acted solely as more recognisable signal of a conflict, Figure 3(b) shows that the lower number of conflicts was reached without an increase in the number of corrections. This result suggests that the display helps in establishing more appropriate spacing between aircraft.



(a) (b)

Figure 3: Effect of the display of time to resolution on prevention of conflicts. Left: the total number of seconds over overlap in landing interval. Right: The number of corrections performed on landed aircraft.

Discussion

The continuation of the learning curve suggest that subjects need more training for a true test of the effect of the display on the planning performance. Further experiments need to be performed. While to CDF may provide a presentation of the occupancy with its relation to uncertainty that is more correct, subjects suggested that the combination of blocks and the PDF may provide a more understandable presentation.

The current approach assumes that aircraft can make an equal adjustment in speed both faster and slower. In reality the available speed change may not be that flexible, and time adjustments may be available through other means (e.g. take-off delay, route adjustment). This would make the available speed envelope much more dependent on individual aircraft. Furthermore, the display did not indicate the cost of speed change, and thus the efficiency of the manoeuvre. A 5 minute delay can be achieved with a very small change in speed (and efficiency) if performed 2 hours before landing, at 30 minutes however, considerable speed changes are required.

The changes in arrival time did not include the cost of a change in schedule to the operator. The cost of a delay depends on the commercial operation of the airline and is therefore unlikely to be equal for each operator or even each flight. The above difference in limitation in available control space and the different cost of adjusting arrival time might need representation in the display to lead to more efficient decisions. As this consists of a number of further constraints on the work domain, the EID framework is expected to be used in this context as well.

The display assumes knowledge on an aircraft's ETA which is available now but might be too uncertain to be of any use. Secondly the display requires information on the uncertainty of the arrival time. The models described used in previous concepts were focussed on the airborne segment of the flight (Whysall, 1998; Mueller et al., 2002; Schaefer, Gizdavu, & Nicholls, 2004). At the horizons considered in this concept, the models will also have to include the predictability of factors on the ground. Therefore, further study will be required to develop appropriate uncertainty models.

Conclusion

This paper demonstrates that it is possible to visualise arrival time uncertainty using the PDF onto the current timeline display. Using such visualisation it might be possible to extend the use of AMAN over a longer time horizon without requiring more accurate arrival time information.

Initial experiments do however show that an interface in which occupancy is presented as a CDF is more difficult to understand for novice users. Especially recognition of available buffer in spacing is more complex in the curved visualisation. To draw more definite conclusions, further experiments, with more training will need to be performed.

References

Barff, A., Favennec, B., Conroy, P., Bellesia, L., Greenwood, J.S., Clark, A., ... Linner, A. (2012). *SESAR P05.06.04 - D28 - Preliminary OSED Ed. 00.01.01*. (Tech. Rep.). SESAR Consortium

Bronsvoort, J., McDonald, G., Paglione, M., Garcia-Avello, C., Bayraktutar, L., & Young, C.M. (2011). Impact of missing longitudinal aircraft intent on descent trajectory prediction. In *30th IEEE/AIAA Digital Avionics Systems Conference*, Seattle, WA. doi: 10.1109/DASC.2011.6096062

Hasevoets, N., & Conroy, P. (2010). *AMAN Status Review 2010* (Tech. Rep.). Brussels: EUROCONTROL.

Hunter, G., (2004). Toward a Standardized Reference Set of Trajectory Modeling Errors: In *AIAA Modelling and Simulation Technologies Conference and Exhibit*. Providence, RI.

Hutchins, E. L., Hollan, J.D. and Norman, D.A. (1985). Direct Manipulation Interfaces. *Human-Computer Interaction,* 1(4), 311–338. doi: 10.1207/s15327051hci0104_2

JPDO. (2008). *NextGen Integrated Work Plan: A Functional Outline*. Retrieved from http://www.jpdo.gov

Mondoloni, S., Paglione, M., & Green, S. (2002). Trajectory Modelling Accuracy for ATM Decision Support Tool. In *International Congress of Aeronautical Sciences (ICAS)*. Toronto, Canada.

Mueller, T., Sorensen, J., & Couluris, G. (2002). Strategic Aircraft Trajectory Prediction Uncertainty and Statistical Sector Traffic Load Modeling. In *AIAA Guidance, Navigation, and Control Conference and Exhibit*. Reston, VA. doi: 10.2514/6.2002-4765

Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions On Systems Man And Cybernetics,* 13(3), 257–266.

Schaefer, D., Gizdavu, A., & Nicholls, D. (2004). The display of uncertainty information on the working position. In *23rd IEEE/AIAA Digital Avionics Systems Conference*, Washington D.C. doi: 10.1109/DASC.2004.1391334

SESAR JU. (2011). *SESAR Concept of Operations at a Glance ED 02.00.00*. (Tech. Rep.) SESAR JU.

Tattersall, A. J., & Foord, P.S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics*, 39(5), 740–8. doi: 10.1080/00140139608964495

Vicente, K.J., & Rasmussen, J. (1992). Ecological interface design: theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics,* 22(4), 589–606. doi: 10.1109/21.156574

Whysall, P. (1998). Future Area Control Tools Support (FACTS). In *USA/Europe Air Traffic Management Research and Development Seminar*, Orlando, FL.

MISMATCHES BETWEEN AUTOMATION AND HUMAN STRATEGIES: AN INVESTIGATION INTO FUTURE AIR TRAFFIC MANAGEMENT DECISION AIDING

Carl A. Westin[1], Clark Borst[2] & Brian Hilburn[1]
[1]*Center for Human Performance Research (CHPR), Voorburg, The Netherlands*
[2]*Delft University of Technology, Delft, The Netherlands*

Future air traffic management will have to rely on more, and more sophisticated, automation to accommodate predicted air traffic. However, studies across various domains have shown that user acceptance of automation decreases when the authority of decision-making automation increases. As a result, low user acceptance could lead to disuse of an automated tool and threaten potential safety and performance benefits. Through a series of human-in-the-loop simulations, the work described in this paper examined the interacting effects of air traffic complexity and strategic conformance, i.e., the fit between human and machine strategies, on automation acceptance in a conflict detection and resolution task. An experiment with 16 professional air traffic controllers showed that strategic conformance is a potentially important construct. That is, conformal resolution advisories were more accepted, led to higher controller agreement, and also reduced response time to proposed advisories.

Future Air Traffic Management (ATM) will have to rely on more, and increasingly sophisticated, automation to accommodate predicted air traffic. This entails automation growth in terms of types of tasks it can perform, and the level of authority and autonomy it can assume, a prediction captured in the SESAR program's definition of five operational Service Levels (SESAR, 2007) intended to guide European ATM evolution.

Studies across various domains have shown that user acceptance of automation decreases when the authority of decision-making automation increases (Kaber & Endsley, 2004; Parasuraman & Riley, 1997). Research has also shown that the predominantly algorithmic approaches used in automation seldom fit well with the more heuristic processes employed by humans. For example, in their exhaustive survey of conflict detection and resolution (CD&R) modeling methods, Kuchar and Yang (2007) concluded that CD&R automation correlates poorly with how controllers prefer to work. Bekier, Molesworth, and Williamson (2012) suggested that there is a "tipping point" for any automated tool, above which controller acceptance of that tool quickly drops. Consequently, a low user acceptance of an automated tool may lead to disuse (and abuse) of that tool (Parasuraman & Riley, 1997) which could severely undermine the intended safety and performance benefits of the tool.

The work in this paper primarily examines the notion of *strategic conformance*, i.e., the extent to which automation's performance and underlying processes are similar to those of the human. Additionally, the interacting effects of strategic conformance with air traffic complexity will also be addressed. Whereas there has been a great deal of empirical and theoretical work into automation in relation to traffic complexity, much less has been done in the area of strategic conformance. Paradoxically, this could in the future become the most critical issue of all, as mismatches between human and machine could threaten initial acceptance of advanced automation. The method applied in this paper for studying human-machine conformance involves presenting controllers unrecognizable replays of either their own solution (i.e., a *conformal solution*) or a colleague's different solution (i.e., *non-conformal*) to a series of pending aircraft separation conflicts. Controllers were instructed that such solutions were generated by automation. Assuming controllers remain consistent in their solution choice over time, this allows us to experimentally manipulate conformance between human and (simulated) machine solutions. Note that by replaying human actions we are in fact simulating automation capable of providing conflict resolutions of the same caliber as a human controller would be able. This replay procedure is inspired by one used many years ago, with which it was shown that operators might be more likely to find faults in automation than in themselves, even when "automation" is an unrecognizable replay of their own performance (Fuld, Liu, & Wickens, 1987).

This paper intends to answer the following questions: are controllers more likely to accept automated advisories when the advisory mimics their own solution? Will acceptance of automation vary with air traffic complexity, regardless of solution conformance? Will workload increase with non-conformal resolution advisories? Does response time vary with conformance of the advisory? First the experimental design is described. The next section presents the results of the study, followed by a discussion and conclusion in the final section.

Experimental Design

We conducted a series of two human-in-the-loop Air Traffic Control (ATC) simulations of increasing automation authority. The first simulation series, considered to be a *prequel* to the main conformance experiment, was designed to capture controllers' manual performance in maintaining safe separations between aircraft. The controllers needed to use an advanced separation assistance tool, the Solution Space Diagram (SSD) developed by Delft University of Technology (Figure 1), to vector aircraft and solve conflicts by issuing speed and/or heading clearances to aircraft (Mercado-Velasco, Mulder & Van Paassen, 2010). In its most succinct form, the SSD is a tactical decision-support tool that visualizes 'go' and 'no-go' areas on a circular heading ring around an aircraft. When vectoring an aircraft into a 'go' area, it will remain free of conflict with all other aircraft and a loss of separation will essentially never occur. In the second simulation series, i.e., the conformance experiment, the controller solutions (and those of their colleagues) to specific conflicts were replayed as automation advisories and plotted within the SSD, after which the controller could either accept or reject the advisory. By plotting the advisory within the SSD controllers could always inspect the validity of the advisory, but also "look around" for better alternatives.

Subjects

Sixteen professional air traffic controllers voluntarily participated at the Shannon Area Control Centre, Ireland. Experience ranged from zero to ten years ($\bar{x}$ = 2.5 years). Twelve controllers currently worked en-route and one controller worked the tower position. Three were students currently undergoing en-route training.

Apparatus

The ATC simulator ran on a portable computer connected to an external 21" monitor. Participants interacted with the simulator through an externally connected computer mouse and keyboard. The ATC simulator was a Java-based application (using OpenGL extensions) that allowed air traffic controllers to control short traffic scenarios. To vector an aircraft, a controller used a computer mouse to click on an aircraft of interest, drag the velocity trend vector to a new conflict-free area on the heading ring (a "clear" area outside the red/yellow areas), and press the ENTER key on a keyboard to implement the vector. Speed clearances (and combined speed and heading clearances) could also be given by using the mouse scroll wheel to either increase or decrease speed. This also increased the radius of the heading band and showed the corrected conflict zones for the new speed settings. This allowed a controller to quickly browse through different speed settings and preview the conflict and conflict-free heading areas for different speeds. Further, no wind conditions were taken into account, all aircraft remained on the same flight level and could not be changed, and the aircraft velocities (and speed clearances) were given in knots Indicated Airspeed (IAS). Further, the aircraft motion was simulated by first order, linear kinematic equations and to keep traffic scenarios sufficiently short and interesting, it was decided to run the simulator four-times faster than real time. Speeding up traffic scenarios in ATC simulators is a common technique to serve this purpose. Finally, the aircraft plots on the display were updated every second to simulate a 1 Hz radar update frequency.

Traffic Scenarios

Each series consisted of 16 traffic scenarios, each based on a squared airspace equal in size (Figure 1). Four baseline scenarios were each rotated in different angles to create three variants, resulting in four scenario groups with four scenario variants in each group. This reduced potential confounding factors, and ensured that initial complexity was the same across scenarios, facilitating comparison between low and high complexity conditions. We aimed to make each traffic scenario repeatable, yet unrecognizable to participants. We maintained sector geometries through scenario rotations in which the relative trajectories and closure angles of aircraft were kept constant, but the entire sector was rotated, and sector entry/exit points renamed.

To guarantee exact replays of controller solutions, each baseline scenario featured only one designed conflict between two aircraft. As a result, any other conflict occurring in a scenario was the consequence of controller intervention. The geometry of the designed conflict was only varied between baseline scenarios. The conflict pair was initially aligned to the exit points and thus required no initial controller interaction. The other aircraft in the sector were considered "noise" aircraft to distract the controller from the conflict pair. Some of the noise aircraft were misaligned with their exit point and displayed in grey, whereas aligned noise aircraft were displayed in green such that the controller could immediately see which aircraft had not yet been cleared to their exit point.

In designing the scenarios it was very important that noise aircraft not interfere with the designed conflict, such that in a replay scenario (i.e., the conformal experiment) the controller would have the same set of solution as

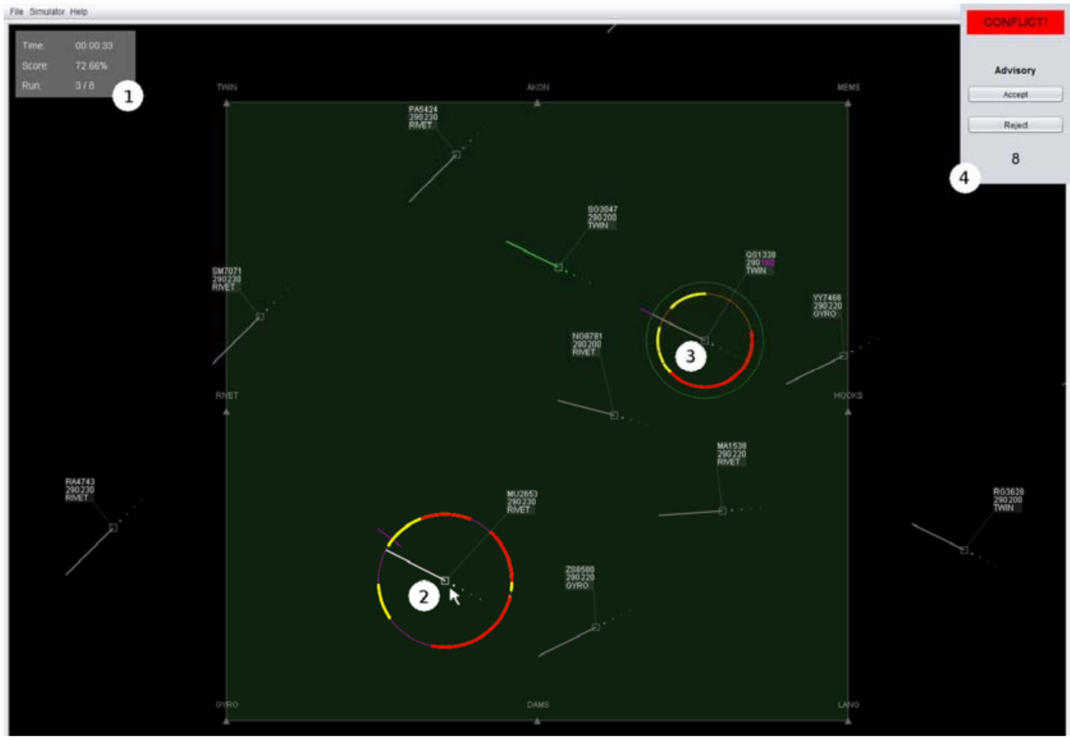


Figure 1. The ATC simulator, showing: 1) the current performance score, time, and runs to go, 2) the SSD of the currently selected aircraft in magenta, 3) the SSD with an automated resolution advisory in an amber color (suggesting a speed decrease to resolve a future conflict), and 4) the advisory dialog window that the controllers could use to either accept or reject the advisory.

observed during the prequel experiment. In the prequel experiment each scenario lasted roughly two minutes. In the conformal experiment each scenario lasted less than one minute in order to reduce the likelihood of significantly changing the traffic situation and thus not being able to guarantee solution replays (i.e., automation advisories) that would solve the conflict. Further, we decided to introduce the automation advisory early in each scenario to prevent participants from solving the designed conflict proactively.

Independent Variables

The experimental followed a within-subjects design with two independent variables, solution conformance (conformal vs. non-conformal) and traffic complexity (low vs. high). A conformal advisory was qualified in terms of aircraft choice, clearance type (e.g., heading change only), and clearance direction (e.g., heading change to the left). A non-conformal solution therefore always featured a different aircraft choice and/or clearance type and direction. Further, non-conformal solutions were derived from solutions provided by other controllers. Complexity was varied through means of aircraft count, and calibrated in a series of test trials. Finally, presentation order of traffic complexity and solution conformance was balanced between participants and traffic scenarios using a Latin Square design.

Dependent Measures

The dependent measures of the conformance study focused on the acceptance of the advisory, the controller's agreement with the advisory (measured on a zero to one-hundred scale), response time, and subjective workload ratings (measured on a zero to one-hundred scale). Response time was measured from presentation of the resolution advisory to pressing the 'accept' or 'reject' button.

Experiment Procedure

The whole experiment took four weeks in total. In the first week, the prequel experiment was conducted with the aim to capture controller resolutions to the designed conflicts. Following briefing and consent procedures, we conducted 16 training runs and 16 measurement runs. Participants were given two main tasks to be performed using the SSD, namely resolving conflicts and clearing aircraft to their intended exit points. A continuously updated performance score reflecting these two task parameters was included to keep participants focused and motivated, and more importantly to prevent scenario recognition and early detection of the designed conflict. To warn the controllers for short-term conflicts, an auditory alert was triggered and the aircraft involved in the conflict were displayed in red.

In the second and third week, prequel data were analyzed and a set of eight conformal (i.e., a replay of controller's own decision) and eight non-conformal (i.e., a replay of a colleague's different but workable) advisories was created for each individual participant. Finally the conformance experiment was conducted in week four. Following a simulator briefing the experiment started with eight training runs, followed by 16 measurement runs. The same scenarios as in the prequel were used but the order varied according to a latin square design. Participants performed the same task as in the prequel experiment, but now were assisted by a higher level of automation that would once in a while provide resolution advisories by proactively auto-select an aircraft in conflict.

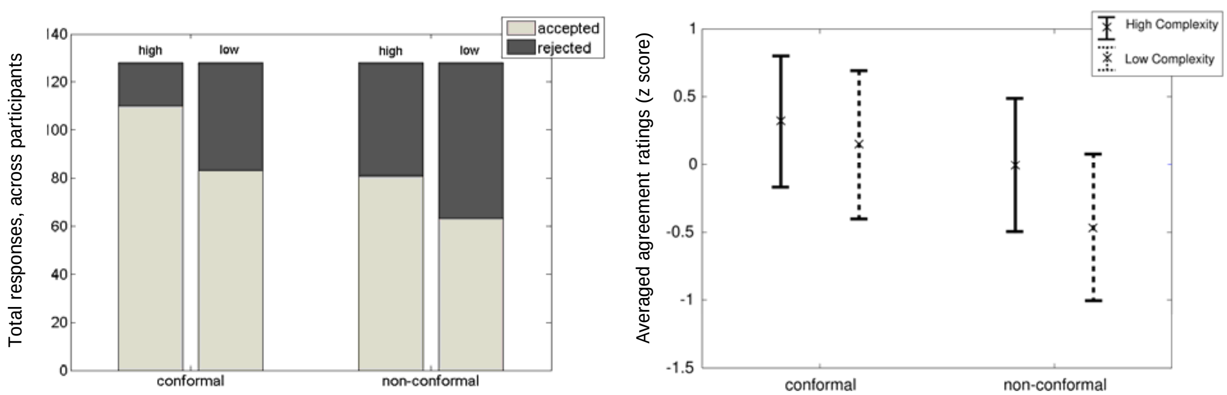
The resolution advisory consisted of either a heading vector, speed vector, or combination thereof, plotted inside the SSD of that aircraft. The resolution advisory was accompanied by a beeping sound and the appearance of a dialog window that the controller had to use to either 'accept' or 'reject' the advisory (see Figure 1). Upon pressing the accept button, the advisory would be automatically implemented. Upon pressing the reject button, the advisory was discarded and the controller had to implement his own workable solution using the SSD. Further, the controllers had 15 seconds to inspect, accept, or reject the advisory. After 15 seconds, the advisory expired and the controllers had to implement their own workable solution. Note that participants were told that an advisory would always solve the conflict, but it would not always suggest the most optimal solution. As such, controllers were encouraged to find better alternatives at their own discretion. After each scenario, participants were given performance feedback in terms of an average performance score. Second, controllers were asked to give ratings on their experienced workload and their agreement with the automation advisory. After the experiment, participants were asked to complete a questionnaire containing information of demographic value and statements (in five-point Likert scale format) querying participant's opinions of the simulator, the SSD interface, and the automated advisories.

Results

Experimenter observations and an analysis of questionnaire data indicated that participants did not recognize scenarios, and more importantly, that the automated advisories were, in half the cases, replays of their own prior conflict resolutions. Questionnaire data also showed that controllers enjoyed the simulator and SSD tool, but did not find scenarios very realistic.

Advisory Acceptance and Agreement with Resolution Advisory

Cumulative accept/reject scores for conformance and complexity can be seen in Figure 2(a). Participants accepted more resolutions advisories in the high complexity scenarios. Comparing conformance, it can be seen that participants accepted more conformal scenarios than non-conformal scenarios. A 2 x 2 repeated measures ANOVA showed a significant main effect for both complexity ($F(1,15) = 11.139$, $p<0.01$) and conformance ($F(1,15) = 10.624$, $p<0.01$) on acceptance. The interaction between complexity and conformance was not significant.



(a) *(b)*

Figure 2. Proportion of advisory acceptances (a) and error bar plots of normalized agreement ratings (b), by complexity and conformance.

Participant ratings of agreement with the resolution advisory revealed a difference between both the complexity levels and conformance levels (Figure 2(b)). The 2 x 2 repeated measures ANOVA revealed that both complexity ($F(1,15) = 7.735$, $p<0.05$) and conformance ($F(1,15) = 18.095$, $p<0.01$) had a significant main effect on the agreement ratings. Agreement with resolution advisories varied positively with increasing complexity and conformal scenarios.

Although a trend was observed in the interaction between complexity and conformance (Figure 3), this was not significant ($F(1,15) = 3.186$, $p>0.05$). This trend suggests that non-conformal solutions tended to be less agreed with than conformal solutions. This effect was especially apparent under low complexity, perhaps because controllers were under less time pressure and would have had more time to evaluate candidate solutions.

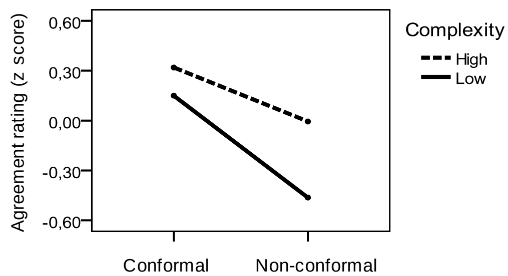


Figure 3. Agreement with resolution advisory ratings (normalized), by complexity and conformance.

Workload and Response Time

The 2 x 2 repeated measures ANOVA showed that workload significantly increased with more complex scenarios ($F(1,15) = 179.950$, $p<0.01$). Neither conformance ($F(1,15) = 0.397$, $p>0.05$), nor the interaction between the two factors was significant ($F(1,15) = 0.266$, $p>0.05$). For response time a main effect was measured by performing a 2 x 2 repeated measures ANOVA, with conformal scenarios having a significant faster response time ($F(1,15) = 9{,}557$, $p<0.01$). Although showing a trend, scenario complexity did not have an effect on response time ($F(1,15) = 4.182$, $p>0.05$). The interaction between complexity and conformance was also not significant.

Discussion

The main objective of this study was to investigate the effect of strategic conformance and complexity on controller performance and acceptance in the context of higher levels of automation. A significant main effect of conformance was observed on acceptance, agreement with resolution advisory, and response time. Conformal resolution advisories were accepted more often, received higher agreement ratings, and were acted upon faster than were non-conformal advisories. Results could, however, not confirm the hypothesis that conformal advisories reduced workload participant workload. These findings suggest that controllers not only discriminate between resolution advisories, but more importantly, prefer resolution advisories that match their own way of working in comparison to advisories generated by their colleagues. That response time increased for non-conformal resolution advisories suggest that controllers questioned the advisories before making the decision to either accept or reject, and that it took longer to conclude that the advisory did not fit with the controllers' own solution. These results support further research initiatives into heuristic-based automation.

Similar to conformance, complexity significantly affected acceptance and agreement with resolution advisory. In high complexity scenarios, acceptance and agreement with resolution advisory was higher than in low complexity scenarios. The lower acceptance and agreement observed in low complexity scenarios could be explained in that controllers had more time to question the advisory and consider more optimal options. The results with low complexity scenarios generating significantly lower workload ratings in comparison with high complexity scenarios, supports this. With increasing workload there are less resources and time available to consider options before making a decision. But these results are also interesting in that they show that controllers disagree with themselves when complexity is lower. It could be indications of learning effects, but it is also a potential indicative that controllers are inconsistent over time, or traits of automation bias.

In this study, controllers were told that all resolution advisories were generated by the automation. How would controllers react if the presumed source was varied between themselves, colleagues, and automation? Further research will attempt to better clarify the concept of automation bias by investigating the effect the presumed source of the advisory would have on acceptance.

It is interesting to see that response time varied with conformance but not complexity, and that workload varied with complexity but not conformance. It could be argued that workload and response time would somehow be connected. However, it is important to underline that response time is applicable to the designed conflict and resolution advisory only, whereas workload considers the entire scenario. The specific workload pertaining to the designed conflict and resolution advisory was not pursued.

The unique experimental design required careful consideration in designing dynamic traffic scenarios allowing exploration of strategic conformance. In order to contain the extent of confounding variables, and increasing experimental control we sought to limit the number of scenarios and through variations of scenario rotation angles increase the measurement span. This experiment showed that it is possible to create "fake" resolution advisories based on participant's own solutions, without participants later recognizing it as their own. It opens up new possibilities in researching not only conformance, but also other aspects such as within-participant consistency and reliability over time.

Despite our best efforts, questionnaire data revealed that controllers did not find scenarios very realistic. This was, however, not surprising considering the limitations of the simulator (i.e., no flight level changes, no wind, limited information, etc.) and that participants were instantly thrown into situations of various complexities. Although we have found some effects of complexity in our study, the results cannot simply be extrapolated to a real operational setting.

Conclusion

These results suggest that strategic conformance is a potentially important construct underlying automation acceptance, and that it can benefit both agreement with automation and response time. A trend toward a conformance and complexity interaction also suggests that the effect on controller agreement with automation advisories is more pronounced under low complexity situations. Though this experiment was considered an important first step, there are several remaining questions to be answered, with respect to both the definition of inherent automation bias, and the nature of such bias as it relates to presumed source of strategic advice.

Acknowledgements

The authors acknowledge the inspiration from EUROCONTROL and the SESAR Joint Undertaking. The work was co-financed by EUROCONTROL on behalf of the SESAR Joint Undertaking in the context of SESAR Work Package E (project MUFASA: Multidimensional Framework for Advanced SESAR Automation). This work reflects only the authors' views and EUROCONTROL is not liable for any use that may be made of the information contained herein. Finally, we would like to thank the staff at the Shannon Area Control Centre, Ireland, for all their amazing support in hosting, facilitating, and participating in the experiment.

References

Bekier, M., Molesworth, B.R.C., & Williamson, A.M. (2012). Tipping point: The narrow path between automation acceptance and rejection in air traffic management. *Safety Science*, 50(2):259-265.

Fuld, R. B., Liu, Y., & Wickens, C.D. (1987). The impact of automation on error detection: some results from a visual discrimination task. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 31(2):156-160.

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.

Kuchar, J.K. & Yang, L.C. (2000). A review of conflict detection and resolution modeling methods. *IEEE Transactions on Intelligent Transportation Systems*, 1(4):179-189.

Mercado Velasco, G.A., Mulder, M., & van Paassen, M.M. (2010). Analysis of air traffic controller workload reduction based on the solution space for the merging task. *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, Toronto, Canada.

Parasuraman, R. & V. Riley (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2): 230-253.

SESAR (2007). Deliverable D3: the ATM target concept, DLM 0612 001 0200a, September 2007.

HUMAN-CENTERED AUTOMATION AS EFFECTIVE WORK DESIGN

Amy R. Pritchett and Karen M. Feigh
Georgia Tech
Atlanta GA, USA.

This paper describes how the challenge of human-centered automation can be recast as the challenge of, first, designing the work performed by a team of agents and then, second, allocating this work amongst all the agents, human and automated, in support of their own needs and capabilities and to foster team goals. The paper starts by formally describing the construct of *work* as a structure which can be formally analyzed and around which other design decisions can be made. It then reviews the requirements of effective function allocation within a team to enable their collective taskwork, and to provide the appropriate teamwork. An example is given that highlights key tradeoffs in designing and allocating work in teams of human and automated agents: no one design can maximize all the desired attributes of human-centered automation.

Work is defined by Webster as "effort directed to some purpose or end." Thus, it is purposeful activity directed at goals established by a concept of operation. Here, we view work as a construct applied at the team level. Further, the notion of work is an ecological perspective: work is achieved by acting on a dynamic environment in response to its demands. This environment can be defined as the aggregation of physical and social/cultural/policy constructs required to describe, constrain, regulate and structure the dynamics of the work; thus, the environment may have inherent dynamics which agent actions need to mirror, may provide affordances which need to be sensed and capitalized upon, and may constrain behavior.

Thus, what overall taskwork needs to happen, and its overall structure and dynamics, is driven by the team goals and by the environment. The team's work emerges out of the collective behavior of all agents in the team, human and automated, even when some of the agents may not see how their activities contribute. The allocation of functions within the team creates the need for additional work: teamwork. This teamwork also requires its own constructs and resources, such that each individual's perception of the environment includes both part of the overall environment and the teamwork aspects created by his/her team members.

From this viewpoint, two things may be designed: the concept of operation defining the goals and structure of the overall task work, and specification of teamwork, including the allocation of functions. The concept of operations is specified at the team level, and establishes core goals for the work; it is constrained by key structures in the environment which the work needs to mirror. The specification of teamwork then brings in the notion of the agents, seeking to allocate functions and identify the constructs within the team that can establish effective human-automation interaction as seen from the humans' perspective.

This paper first summarizes the construct of work as a structure which can be formally analyzed and around which other design decisions can be made. It then reviews the requirements of effective function allocation within a team to enable their collective taskwork, and to provide the appropriate teamwork. An example is given that highlights key tradeoffs in designing and allocating work in teams of human and automated agents: no one design can maximize all the desired attributes of human-centered automation.

Modeling Work

Despite a common tendency to focus on technology design, designing the work can be the more important concern in establishing effective human-automation interaction. Indeed, work design is a harder task than technology design, as specifications of work, such as concepts of operation, must intrinsically integrate the economic and safety metrics by which the total system will be evaluated, the potential contributions of (or constraints on) technology and human performance, and the regulatory, policy and procedural considerations in allowing access to - and defining interaction within - the collaborative functioning of the system.

Thus, the foundation of human-centered automation is laid in the work design. At such an early stage, human in the loop evaluation is not possible – the training, procedures and technology are only specified in terms of the functions required. Instead, the important construct is: *If everything, and everyone, in the system performs their functions perfectly, what will emerge?* The answer is created by the interplay of the work environment (as defined by physics and regulations) and the team acting upon the environment. Concepts of operation can be constructed poorly when they are sensitive to small variations in how the work is performed, or where they assume actions will be performed with a speed or detail that is not possible or for which information is not available. For example, air traffic concepts of operation applying optimized profile descents must create the work activities that regulate the physics of an aircraft descending in a fuel-efficient manner, while recognizing that key variables – the aircraft performance, the wind profile through the descent, constraints on the aircraft to fit within the traffic stream – are known (or only partly known) at different locations and at different times, yet key decisions to descend earlier or later can have profound impacts on the aircraft's ability to follow its individually-optimized descent and fit within the broader traffic flow.

To ensure the overall specification of work is sound, work activities can be first modeled in detail without requiring detailed models of the agents who may perform them. Conceptually, this analysis is best conducted with simple models of human performance such that any problems can be clearly linked to the feasibility of a concept of operation. Further, once the broader dynamics of the work are established, a concept of operation can be examined for its robustness and resilience: *What if something doesn't go perfectly?* Here, the system's response to unexpected events can be modeled and simulated. These unexpected events may stem from several sources: exogenous inputs to the system (for example, on an air traffic system an unexpected tailwind or thunderstorm); technology (for example, the failure of a radar system); or from human performance (for example, limit on the number of simultaneous activities that can be performed). The work involved in responding to these events will be emergent and dynamic, and a concept of operation can be designed to be more (or less) robust and resilient).

Work can be analyzed in several ways. Approaches such as Contextual Design (Beyer & Holtzblatt, 1988) and Cognitive Work Analysis (Bisantz and Roth, 2007) provide qualitative and visual presentations of the work that are intended to guide and inform designers. Our own recent efforts have established a computational framework that enables work to be computationally modeled and simulated early in design (Pritchett, 2013), first to support analysis of the concept of operation (i.e. the required taskwork of the entire team) and then to examine the design of the team itself (i.e. allocation of functions within the team, and their teamwork).

Requirements for Effective Function Allocation

Function allocation distributes work between agents, human and automated, within a team. The following requirements for effective function allocation can be noted from first principles and the literature. The following summaries are a review from more extensive discussions by Feigh, Pritchett and Kim (in review).

Requirement 1: Each agent must be allocated functions that it is capable of performing.

Every agent in the team must be capable of each of the functions assigned to him/her/it, viewing each function in isolation. In a very coarse sense, such a strategy is supported by assessments of what "Men Are Better At" to what "Machines Are Better At." From this perspective, automation can serve to provide functions that a human cannot perform at all or with sufficient reliability. However, the automation must not be brittle such that, when placed outside its boundary conditions, such automation appears to its operator to fail. *Thus, a prediction of whether the automation will be placed outside its boundary conditions is itself a valuable metric that implies potential concerns with the resilient performance of the team.*

A further consideration in creating effective human-automation interaction examines responsibility and authority. Except when automation is proven to provide safety in all foreseeable operating conditions, humans remain vested with the responsibility for the outcome of automation's actions, a situation termed the "responsibility-authority double-bind" (Woods, 1985). If the human cannot knowledgably oversee the automation, they are forced to 'trust' the automation. However, without a concrete basis for assessing if the automation is correct, humans often over- and under-trust the automation (Parasuraman & Riley, 1997); either way, incorrect trust is viewed as human error, despite its basis in the function allocation. *Thus, identification of mismatches between responsibility and authority is itself a valuable metric that implies potential concerns with trust and reliance, and that requires monitoring by the human.*

Requirement 2: Each agent must be capable of performing its collective set of functions.

The metric for success for this requirement is whether each agent can perform his/her/its collective set of functions under realistic operating conditions. *Thus, prediction of the taskload placed on the human operators – or, where possible, workload experienced by the human operators – is a valuable metric of function allocations.* To fully address known issues with taskload corresponding to human-automation function allocation, such assessments must consider the full range of activities required, including underlying cognitive activities around information gathering and judgment, and requirements to monitor automation, in addition to explicit manual activities. Further, metrics of workload should consider not only aggregate or average workload, but also workload spikes and periods of complacency.

Further, human-centered automation requires that the function allocation establish coherent roles for agents. One attribute of a coherent function allocation can be viewed from the bottom up – within each agent its functions share (and build upon) obvious, common constructs underlying all their activities, such as a shared information and knowledge basis, and the allocation prevents conflicts between the actions of different agents. Another attribute can be viewed from the top down - the functions collectively contribute towards work goals in a manner that is not only apparent to the human, but that can be purposefully coordinated and adapted in response to context. Thus, the coherence of the functions allocated to each human is itself intrinsically an important construct warranting its own analysis.

Requirement 3: The function allocation must be realizable with reasonable teamwork.

Each different function allocation of the same taskwork demands its own unique set of teamwork functions, including functions for human-automation interaction and for human-human coordination. The impact of this teamwork must then be considered from the perspective of the previous two concepts - can each agent perform each of his/her/its teamwork activities in isolation, and can each agent perform its assigned set of both task work and teamwork functions?

Members of good teams are able to anticipate each other's information needs and provide information at useful, non-interruptive times. However, too often automation is 'clumsy:' it unduly interrupts its human team members because, whereas humans can implicitly sense information about whether other team members would benefit from an interruption, automation historically cannot. *Thus, the potential for a function allocation to cause agents to interrupt each other is an important construct to be analyzed.* In some cases, such as poorly-timed output from automation, such interruptions may be unwarranted; in other cases, different function allocations may require agents to interrupt each other more or less depending on how their functions are allocated and, perhaps, inter-leaved.

Requirement 4: The function allocation must support the dynamics of the work.

Analysis of a function allocation should identify situations where, for example, the interleaving of functions assigned to disparate agents requires significant co-ordination or idling as one waits on another, or where workload may accumulate, or where one agent will be unduly interrupting another, or where executing prescribed procedures may conflict with other work demands, or where automation may be placed outside its boundary conditions. These issues were discussed in the preceding sections, but are repeated here to note their dynamic nature.

Further, resilience is fostered when a human agent may select strategies (courses of action) appropriate to the state of the environment and their own capabilities. The ability of each human in the team to adapt to immediate context has been found to reflect a good balance between the demands on the human and the resources available to them in terms of information, knowledge and time available (Feigh & Pritchett, 2006). However, such adaptation can be constrained or eliminated by an overly prescribed (or proscribed) function allocation, particularly where human-automation interaction dictates a specific sequence of activities from the human. The adverse effects of such overly prescribed function allocations have been found to manifest in work-arounds or dis-use of automation (Feigh & Pritchett, 2010; Parasuraman & Riley, 1997). *Thus, the ability to which a function allocation can accommodate a reasonable variety of human adaptations to context should also be analyzed and fostered.*

Likewise, human-centered automation should foster the humans' ability to maintain a stable work environment. A function allocation may aggravate inherent environmental unpredictability by, for example, limiting human agents' ability to view important aspects of the environment or by distributing functions in a way such that one agent will trigger the requirement for another to act. In addition, a trade-off exists when designing function allocations between maintaining predictability vs. dynamically allocating functions (Miller & Parasuraman, 2007). *Thus, humans' ability to predict their activities has intrinsic value and should be fostered*.

Requirement 5: The function allocation should be the result of deliberate design decisions.

Changes in operational concepts may be incremental and constrained by current-day technologies, procedures, personnel and/or policies; in other cases, changes in concepts of operation may represent significant innovations in which constructs such as common work practices and relationships between tasks and tools must be significantly altered. Either way, designers need to simultaneously consider the economic and safety metrics by which the total system will be evaluated, the potential contributions of (or constraints on) technology and human performance, and regulatory, policy and procedural considerations. *Thus, the design of human-centered automation should consider not only each agent's experience, but also simultaneously consider the cost and performance of the combined efforts of the human-automated team.*

Conclusion: Perfect Human-Centered Automation is Impossible

In an earlier study we examined four function allocations using computational simulations of work, ranging from full autoflight with datalink (FA1) through progressively 'less automated' conditions to pilot control of the trajectory by setting immediate autopilot targets (FA4) (see Feigh, Pritchett and Kim, in review). In these simulations we also assumed that the human agent (the pilot in this case) might exhibit three different behaviors, as represented by the Opportunistic, Tactical and Strategic cognitive control modes (CCM).

Figure 1 reflects a subset of the metrics collected to examine the ability of the concept of operation and function allocation to meet the requirements noted above and to meet the mission goals as measured by metrics such as time to land. In this figure, the metrics are normalized such that 100% represents the ideal: perfect human-centered automation would have 100% on each of these metrics. Instead, each function allocation scores higher on some metrics and lower on others. The more automated function allocations required better (less) interaction with the pilot but were less predictable to the pilot, made for a lower coherency role for the pilot and interrupted the pilot more. The less automated function allocations provided a more coherent role for the pilot and more predictability, at the expense of requiring them to do more of the work. Further, all of the function allocations assumed the pilot would perform monitoring activities that we predict the pilot would shed in the opportunistic and tactical CCM.

In the end, all of the function allocations met the mission goals in this case. This reflects a situation common in aviation – the agents can adapt and respond to the environment to get things done. The challenge in designing human-centered automation is identifying how to design the work – the concept of operation and the function allocation within it – that strikes the right balance between key trade-offs inherent to divvying up the work to reduce workload, yet maintain coherency, predictability and reduce interruptions.

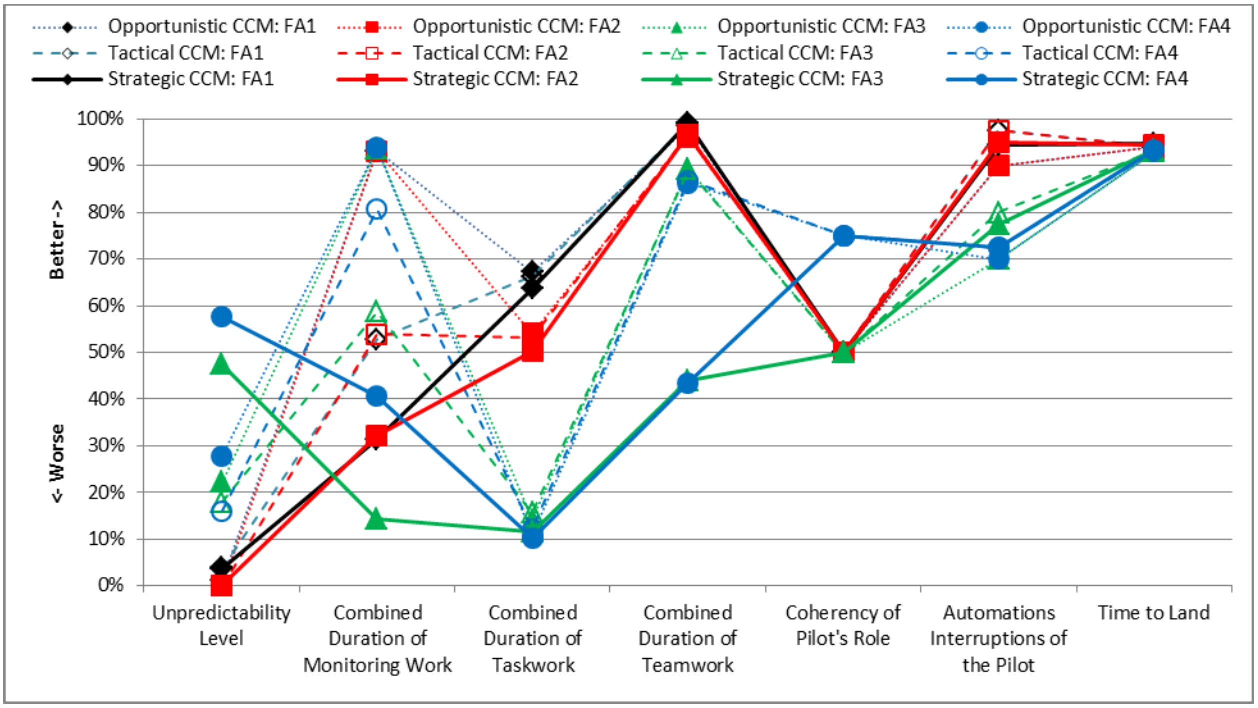


Figure 1. Key metrics of four function allocations between pilot and autoflight system, from the most automated FA1 to the least automated FA4, in conditions where the pilot behavior follows opportunistic, tactical and strategic cognitive control modes (CCM).

References

Beyer, H., & Holtzblatt, K. (1998). *Contextual Design*. San Diego, CA: Academic Press.

Bisantz, A.N., & Roth, E.M. (2007). Analysis of cognitive work. In D.A. Boehm-Davis, (Ed.), *Reviews of human factors and ergonomics, Vol. 3* (pp. 3–35). Santa Monica, CA: Human Factors and Ergonomics Society.

Feigh, K.M., & Pritchett, A. (2006, October). Contextual control modes during an airline rescheduling task. In *Annual Meeting of the Human Factors and Ergonomics Society*, San Francisco, CA. Santa Monica, CA: Human Factors and Ergonomics Society.

Feigh, K.M., & Pritchett, A.R. (2010). Modeling work for cognitive work support system design in operational control centers. *Journal of Cognitive Engineering and Decision Making*, *4*, 1–26.

Feigh, K.M., Pritchett, A.R. & Kim, S.Y. (in review) Modeling human-automation function allocation. Submitted to *Journal of Cognitive Engineering and Decision Making.*

Miller, C.A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors*, *49*, 57–75.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*, 230–253.

Pritchett, A. (2013). Simulation to assess safety in complex work environments. In J.D. Lee and A. Kirlik (Eds.), *The Oxford Handbook of Cognitive Engineering.* Oxford University Press.

Woods, D. (1985). Cognitive technologies: The design of joint human-machine cognitive systems. *AI Magazine*, *6*(4), 86-92.

VALIDATING A MODEL OF AUTOMATION SUPPORTING THE ROBOTIC ARM CONTROLLER

Christopher D. Wickens, Angelia Sebok, A. Marc Gacy
Alion Science and Technology, Boulder, Colorado
Huiyang Li and Nadine Sarter
University of Michigan, Ann Arbor, Michigan

A simulation of the space robotic arm navigation task is described. This simulation is used in both a human-in-the-loop simulation experiment to generate human performance data, and is coupled with a computational model of the human: MORIS, whose outputs are compared to the human operator data for both nominal conditions at three levels of operation, and for automation-failure conditions. Scan mediated model predictions of automation failure response are validated by the human performance data.

This paper describes an effort to model the astronaut controller of a space-based robotic arm, such as that found on the space shuttle or International Space Station. Such an arm is designed to both grasp objects within its “hand” (called the end-effector or EE) and transport them to 3D points in the environment by manipulating arm joints of the shoulder elbow and wrist, along multiple degrees of freedom. Two 2-axis controllers are typically employed; one controlling rotation of the wrist and EE, and the other controlling 3D translation of the EE.

While carrying out the 3D navigation, the human operator must continuously be aware of constraints on the shoulder, elbow and wrist rotations to avoid what are called “singularity lockups” that freeze the arm in place, following which a time consuming recovery process is required. Naturally the operator must also monitor the arm in the workspace to avoid collision of wrist and elbow with hard constraints (e.g., obstructions) in the space.

As shown in figure 1, the operator monitors and controls the 4D trajectory (XYZ and time) through any of 6 cameras (two depicted in the figure), viewable through 3 different “viewports” or monitors, selecting at any time, those cameras that provide the best spatial understanding of the arm and EE relative to hazards and target destinations. The operator can also monitor joint angles on a separate display to assess their proximity to singularities and other abnormal states. A typical arm mission can be described in 3 phases: initiation of the appropriate movement; movement itself, and a final alignment and grasping (or releasing) of the payload by (from) the EE. Figure 1 presents a schematic layout of the workstation & workspace.

Many aspects of this task are analogous to the aircraft pilot, flying a 3 phase trajectory (departure, cruise, approach) while both navigating, and also preserving stability, with information provided by multiple displays. The manual operation of both flying and robotic arm manipulation can impose extremely high levels of cognitive and motor workload. For the robotics operator, this can be moderated by slowing or pausing the operation. However for the aircraft pilot, this workload has been mitigated by several layers of automation (Ferris, Sarter & Wickens, 2011). In particular, relevant to the current project we consider automation of guidance, via displayed vectors (e.g., a recommended flight path, much like the highway in the sky (HITS) display [Prinzel & Wickens, 1999]); and automation of control, akin to the coupled autopilot in the FMS, where trajectories can be flown merely by specifying XYZ endpoints. In both cases, reductions in workload and flight path error have been achieved, although in the case of the autopilot, the reduced workload comes at a cost of reduced situation awareness.

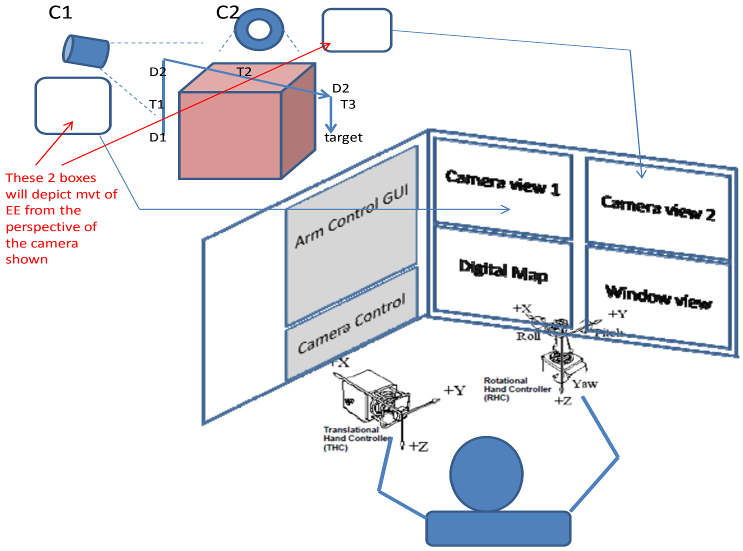


Figure 1. Schematic description of the robotic arm workspace (upper left) and displays. The figure depicts only two (rather than 6) camera views (C1 & C2) and a window view, and, within the workspace, a typical mission, to move the EE up, across a table and down to a target.

In marked contrast to the pilot, the robotic arm controller has not been well supported by corresponding layers of automation; and so our project has focused on development of automation support. This has been imposed in three ways. For **trajectory control**, we have created analogies to the two forms described for aviation: automation guidance, achieved by a 3D trajectory line through the workspace displayed on the camera view monitors, and full automation (autopilot) control. For hazard control, we have implemented an automatic collision warning system. For **camera control**, we have implemented intelligent guidance for the optimal camera view. Trajectory automation however will be the focus of the current paper.

We note that the two levels of trajectory automation (auto-guidance and auto-control) correspond closely to the levels of automation within the stages-levels taxonomy proposed by Parasuraman et al (2000) as an extension of the Sheridan and Verplank (1978) **levels of automation** scale. This distinction becomes of paramount importance, given the well-validated finding that the higher the level of automation, the better it works during normal operations and the lower the workload; but the **greater are the penalties when automation fails**. Onnasch, Wickens, Li and Manzey (submitted) carried out a meta-analysis of automation failure studies to validate this relationship, which they referred to as the "lumberjack analogy": like trees, *the higher they are, the harder they fall*. Thus one component of both our human-in-the-loop (HITL) simulation experiment, and our cognitive model predictions will examine this relationship between normal and failure performance.

Methods and Modeling

The overall content of this research endeavor included two parallel but interacting efforts: (1) At the University of Michigan, we developed a computer simulation of the robotic arm itself, modified from the original specifications of the system used to train astronauts at NASA, a simulation called BORIS. We employed this to gather HITL data of 36 well-trained subjects, operating the simulated arm under both nominal, and unexpected "failure" conditions, along a 3 segment trajectory that required multi-axis control and avoidance of a table hazard in the middle of the workspace (see figure 1). Details of these results are provided in Li et al (submitted). (2) At Alion Science, we developed a computer simulation model of the robotics operator using BORIS, a simulation which we called MORIS. The architecture, parameterization and validation of MORIS will be the focus of the current paper.

MORIS contains four linked sub-models, as shown in figure 2. At the left is a utility-based **decision model**, that decides, based upon maximum utility and pre-established rules, which modes to select, which trajectory to select and which cameras to choose for the two viewports. Input to the camera selection decision are outputs from the **FORT model** (Frame of Reference Transformation), which continuously computes the cognitive load of translating a given camera view into a control action (Wickens, Keller & Small, 2010). This model assigns penalties to the extent that a given view is closer to parallel to the line of sight into the display (McGreevy & Ellis, 1986; Wickens, Vincow & Yeh, 2005), and to the extent that the view provides EE motion information that is **incompatible** with the direction of control motion, or is hampered by poor visibility. In addition to deciding which camera to choose, the ubiquitous FORT model also influences the value or utility of each camera view to visual attention (via the SEEV model shown at the top of the figure, as discussed later), the fluency of control in the trajectory model, and provides an input to the perceived workload output of the model. The **trajectory model**, influences the fluency of actual control, and is heavily influenced by automation level (see below). Finally, a **visual attention model (SEEV**; Wickens, 2013; Wickens et al, 2003), predicts the pattern of eye movements across the 6 displays, based in particular on the **effort** to move attention, and the **expectancy** (bandwidth) and **value** of changes within each display (EEV within SEEV). The latter parameter is heavily determined by task relevance and the FORT based utility of each camera view of each display. As discussed below, expectancy and value are influenced by automation level

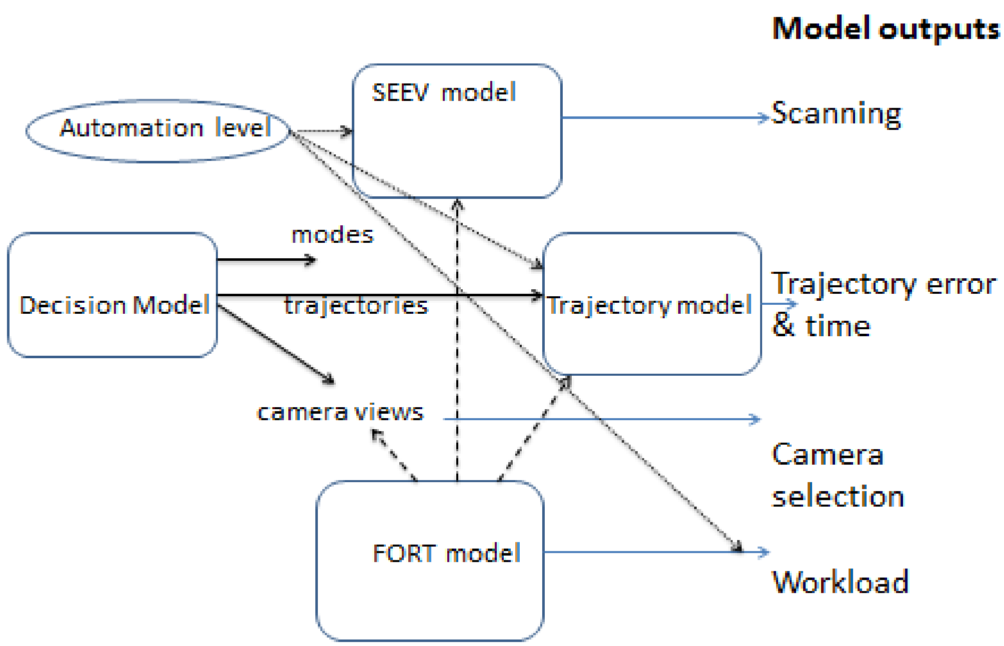


***Figure 2*:** Architecture of MORIS. Dashed lines represent FORT influences. Light blue lines show how model outputs are generated**.** The influences of automation level are dotted arrows.

Implicit in the model are three critical assumptions regarding automation and its linkage to SEEV parameters: (1) The highest level of automation (autocontrol or AC) directly diminishes the SEEV **value** of the camera monitors to the task of trajectory control (e.g., the relevance of those views to the navigational task). (2) Also in SEEV, the bandwidth or expectancy of the displays portraying EE motion is directly proportional to the tracking error observed from the empirical data at Michigan. Thus we assume that the higher error observed in the less automated conditions results from more frequent, and less stable corrections and requires more frequent sampling of the EE displays (Moray, 1986; Wickens, Goh et al., 2003). (3) We identified channel specific workload differences across the three trajectory automation conditions. Specifically, we predicted that *cognitive workload* in the autoguidance condition would be half the value in the manual condition. This is because in the manual condition, subjects needed to envision the correct trajectory between XYZ waypoints, whereas in the autoguidance condition these waypoints were directly displayed. Motor workload was little changed between the two conditions, as both required manual correction. In contrast, motor workload was assumed to be zero for the autocontrol condition, as neither response selection nor execution was required.

Results

Table 1 contains key aspects of the data from the Michigan simulation in the left of each column, that represent both the target values which we used to set parameters for the MORIS model as well as, for the two failure trials that occurred at the end of the experiment, targets for failure response validation, as discussed below. In the case of every dependent variable in the table, differences across the three levels were statistically significant ($p<.05$; see Li, Wickens et al., submitted). Presented in *italics* in the right side of the cells are the MORIS model predicted values, discussed below.

Table 1. Empirical HITL simulation data (left side of each cell) and MORIS model output (right side), as a function of trajectory automation level. Model data are not yet available for all measures.

Variable	Manual		Auto-guidance (AG)		Autocontrol (AC)	
Completion time (s)	440	*440*	401	*413*	215	*334*
Trajectory error	81		14		1.0	
Camera switches #	15.2		10.9		7.0	
Subjective workload	4.6	*4.6*	3.4	*3.8*	2.6	*3.0*
Trial 6 failure			**113**	***.40***	**144**	***.55***
Trial 7failure	**9/24 = .37**	***.37***	**8/24=.33**	***.40***	**20/24 =.83**	***.55***

(a) Parameterization

- Completion time is expressed as a value normalized to the maximum empirical value (manual condition). The correlation between obtained and observed completion time is 0.99
- The trajectory model simulated **tracking error** by adopting a threshold error, above which MORIS generated a linear closed loop correction to any path away from the target path. We assumed that this threshold was five times greater for the manual than the autoguidance condition, since in the manual case, the target path needed to be imagined or envisioned, whereas in the autoguided, it was directly visible. This is the basis of prediction of the two error measures for manual and AG. The autocontrol error was assumed to be close to 0, with a near perfect autopilot.
- The predicted number of camera switches was simulated based upon an algorithm in which, whenever computed FORT penalties for both camera views exceeded a key threshold, the currently unviewed camera with the lowest FORT penalty would be selected.
- Subjective workload was simulated by summing the predicted workload across (visual, cognitive, and motor) channels; based in large part upon the differing automation demands described above. However cognitive workload was also augmented by higher FORT values returned by the camera views across all three conditions, and by visual scanning. As with completion time, we normalized to the maximum (manual) value. The correlation between predicted and obtained workload measures was 0.99

(b) Validation: Automation Failure trials

- **Failure: trial 6**. On trial 6, a failure unique to the two automation conditions was imposed by depicting the guidance line (coupled, in the AC condition, with the actual flight path chosen by automation) along a path different from the correct direction to the final target on the third segment. Our measure of the ability to detect and correct this errant automation was the size of deviation from the correct path to the target.

- **Failure: trial 7**. On the final trial, for all three automation conditions, a proximity warning alert that had functioned correctly on all previous trials (including training) now failed, by remaining silent even as proximity limits were violated. This violation was guaranteed by directing the EE guidance too close to the table in the two automated conditions, and by providing XYZ coordinates for a corner turn that would yield a similar proximity violation trajectory in the manual condition. Our performance measure was a pooled measure combining the number of violations of proximity limits with the number of actual collisions with the table or wall hazard, both summed across subjects for this single trial.

We observe in the empirical data of both trials 6 and 7, a marked decrease in performance at the higher (AC) compared to the lower (AG) level of automation, consistent with the lumberjack analogy, and the better performance and reduced workload at the higher level, seen by measures in the upper rows of the table. We also note however that autoguidance automation does **not** induce poorer performance than manual automation on failure trial 7.

While the degree of model fit to the empirical data for the normal trials in the two upper rows for which model outputs were available was, to some extent expected, since we used those data to essentially "parameterize" the model, the same cannot be said for the failure trials. Here we made some basic assumptions grounded in eye movements and based upon "complacency theory" in human automation-interaction (Parasuraman & Metzger, 2007, Wickens, Dixon, Goh & Hammer, 2005). These assumptions allowed us to predict scanning behavior during normal trials on the basis of the SEEV model and use these to **infer** the manner in which automation-induced differences in scanning across the three conditions, would modulate fault detection ability. More specifically we assumed that (1) following the programmed failure deviation, ***a violation would occur if the trajectory was not manually corrected within 3 seconds***, and (2) ***complacency-induced scans*** away from the camera window where such deviation would be visually apparent, ***left that now-neglected area unattended***; hence this would create a human failure to notice the automation failure, if the eye did not return there before 3 seconds had elapsed. SEEV provided scan data, and in the SEEV model in the manual condition both expectancy and value were set to their maximum level, as described previously. In the autoguidance condition, expectancy was 1/3 maximum reflecting the large decrease in tracking error (see table 1) but value was retained at near its maximum level. In the auto control condition, both parameters were set to minimum. The SEEV scanning data and a noticing model (NSEEV) provided the probability of miss data shown in italics in the trial 7 failure trials; values that very closely approximated the obtained data, and a correlation of r=0.97 was obtained between model predicted and human generated data.

Discussion

In this paper we have presented the development of a computational computer simulation model of the human robotic arm controller. To our knowledge this is the first such effort. The model contains four submodels of spatial transformations, visual attention, decision making and trajectory control. For four outputs of the model, trajectory time and error, camera selection and workload, there was no a-priori basis for selecting parameters that would fit the experimental data from the HITL simulation of the arm controller. Hence agreement between predicted and obtained values was to be expected. However for one particular aspect, off-nominal automation failure response as a function of the level of automation in trajectory control, our effort produced something closer to a true (and successful) validation. We made a-priori assumptions of how level of automation would influence visual scanning (complacency) to critical areas where the automation failure would be noticeable, hence predicting failure response fluency. These model predictions were well validated. Additional empirical data will be sought to continue validation.

References

Ferris, T., Sarter, N., & Wickens, C. (2010). Cockpit automation: still struggling to keep up. In E. Salas and D. Maurino (Eds). Human Factors in Automation, 2nd ed. Amsterdam NL. Elsivier.

Li, H, Sarter, N., Wickens, C. & Sebok, A. (2013 submitted) Supporting Human-Automation Collaboration through Dynamic Function Allocation: The Case of Space Teleoperation. Submitted to Annual Proceedings Human Factors & Ergonomics Society.

Li, H., Wickens, C., Sarter, N., & Sebok, A. (submitted) Types and Levels of Automation in Support of Space Teleoperations. *Human Factors Journal*

Metzger, U., & Parasuraman, R. (2005). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. Human Factors, 47, 35–49.

McGreevy, M. & Ellis, S., (1986)The effect of perspective geometry on judged direction in spatial information instruments. *Human Factors, 40*, 443-451.

Moray, N. (1986). Monitoring behavior and supervisory control. In L. K. K. R. Boff, & J. P. Thomas (Ed.), Handbook of perception and performance, Vol. II New York: Wiley & Sons.

Onasch, L., Wickens, C., Li, H. & Manzey, D. (submitted) Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis. Human Factors Journal.

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 – Part A: Systems and Humans, 30, 286-297.

Prinzel, L. & Wickens, C.D. (2009) Preface: special issue on Synthetic Vision Systems. International Journal of Aviation Psychology, 19, 1-9

Sheridan, T. B., & Verplank, W. L. (1978). Human and computer control of undersea teleoperators. (Technical Report). Cambridge, MA: Man Machine Systems Laboratory, MIT.

Wickens, C.D (2013). Noticing events in the visual workplace: The SEEV and NSEEV models. In R. Hoffman & R. Parasuraman (Eds). Handbook of Applied Perception. Cambridge, U.K.: Cambridge University Press.

Wickens, C.D., Goh, J., Helleberg, J, Horrey, W., & Talleur, D.A. (2003). Attentional models of multi-task pilot performance using advanced display technology. Human Factors, 45(3), 360-380.

Wickens, C. D., Vincow, M., & Yeh, M. (2005). Design applications of visual spatial thinking: The importance of frame of reference. In A. Miyaki & P. Shah (Eds.), Handbook of visual spatial thinking. Oxford University Press.

Acknowledgments

This study was supported by a grant from the NASA Johnson Space Center Human Research Program (NNX09AM81G). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of NASA. The authors wish to thank technical monitors Barbara J. Woolford and Douglas Wong for their support of and interest in this work.

PILOT DECISION MAKING: MODELING CHOICES IN GO-AROUND SITUATIONS

Andreas Haslbeck, Armin Eichinger, Klaus Bengler
Institute of Ergonomics, Technische Universität München
Munich, Germany

Pilot decision making is highly influenced by cockpit information displays. Decision quality could benefit from knowledge of temporal and individual influences on decision making under time pressure that suggests leverage points for cockpit or process design. In a recent flight simulator experiment, airline pilots were presented a realistic landing scenario. During the approach phase, instruments indicated weather conditions suggesting a go-around decision to be taken. The alternative decision consists of landing in spite of illegitimate strong tailwind. Gaze tracking analysis identified, whether relevant display information was picked up by the pilots. The time between checking the aircraft's wind indicator and the moment of decision was taken as predictor of choice to go-around. Modeling of pilots' choice behavior shows strong influences of the predictor analyzed. A comparison of long-haul captains and short-haul first officers shows dependency of decision-behavior on level of practice and training.

Introduction

One important aspect of good airmanship is pilots' decision making (FAA, 2004; DeMaria, 2006). A pilot's ability to soundly decide duly prevents hazardous situations. While several aspects of good airmanship like manual flying skills can be taught and exercised at flight school, long-term experience is needed to build up comprehensive knowledge for an aviator to find appropriate decisions in a certain situation. One potentially hazardous situation is the approach phase, representing more than one third of all fatal accidents (IATA, 2011; Boeing, 2012). Two typical accident categories defined by the International Air Transport Association are runway excursions (23% of IATA listed aircraft accidents in 2010) and hard landing (5%). In-depth analysis has shown, that in 35% of the runway excursions in 2010, meteorology has been a contributing factor. To complement this information, in one fourth of all cases, the flight crew has failed to go-around after an unstabilized approach (IATA, 2011). The safety reports of the years before have shown very similar numbers and evidence. One lesson to be learned from these reports is that a go-around can be a safe decision to master the high-risk situation of a hazardous approach.

Taxonomy of go-around behavior

The focus of this experiment is the pilots' behavior in an approach scenario, where a go-around has to be performed by the pilot flying (PF) because of an illegitimate high tailwind (Table 1). The PF should be aware of this wind situation and trigger the go-around by himself (type 1). If he is not aware of the tailwind, a cue by the pilot monitoring (PM) can lead the PF to trigger the go-around (type 2). In both cases it may happen that the PF does not trigger the go-around because of a decision to land in spite of the tailwind (type 3) or because of not being aware of the wind even when a cue is given (type 4).

Table 1.

Different types of pilots' behavior concerning the decision of a go-around.

	pilot is aware of wind situation	pilot is not aware of wind situation
pilot is going around (PGA)	Type 1	Type 2
pilot is not going around (PnGA)	Type 3	Type 4

For type 1 und 2 the time intervals between different wind checks can be calculated (Figure 1). If a pilot is aware of a wind potentially differing from the ATC information, he should early perform a first wind check (t_1) in the final approach (below 1,000 ft above ground level) and should repeat this check continuously until a final decision to (not) go around is made. The final wind check before the go-around is also measured (t_2). If only one wind check is performed first and last check time coincide (t_1=t_2).

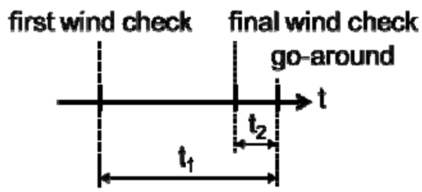


Figure 1. Times measured between different wind checks.

As third gaze indicator the time duration between first (t_1) and final (t_2) wind check (t_1-t_2) can be calculated.

Research Questions

One research question of this study is: Does the amount of experience influence pilots‘ decision making behavior? In a first analysis two groups of pilots that differ in their level of training are compared in regard of frequencies of go-around behavior as described in Table 1. In a second analysis, this behavior is further detailed independently from the two different groups in terms of related underlying mechanisms?

Dual-process influences on pilot decision making. To understand the mechanisms underlying these go-around decisions, we put them in context of established behavioral taxonomies and propose continuous usage of relevant information displays as an important predictor of decision making. Many process-models of pilot decision making in a first step assess the situation by observing information and data scanning (e.g. Jenkins et al., 2008; Orasanu, 1995) and branch out depending on the interpretation of this assessment. A central source of decision errors is thereby made explicit: the lack of consideration of important data displays; a perceptual step that also builds the foundational level of Endsley's conception of situation awareness (e. g. Endsley, 2006).

What are the driving forces for pilots to consider relevant information sources, i.e. important data displays? Rasmussen's classification of action identifying skill-based, rule-based, and knowledge-based behavior can help to localize relevant mechanisms (Rasmussen, 1983). As designers of man-machine interfaces we intend to facilitate behavior that is situated near the lower, skill-based part of the taxonomy. The reason for this is illuminated when put in context of dual-process theories of thinking and decision making (e. g. Evans, 2008; Kahneman & Fredricks, 2002; Kahneman, 2003, 2011). Skill-based behavior is a function of system 1 whose processes are automatic, opaque, and effortless (Kahneman, 2011). System 2 is a highly flexible regulatory entity with potential control of system 1 suggestions for action. Its processes are slow, self-aware, and effortful. Behavior on a skill-based level requires less effort and induces less workload than rule- or knowledge-based behavior that is in the domain of system 2. From an energetic self-regulatory perspective that leads to the tendency to invest not more effort than is required in a task. System 2 usually endorses system 1 suggestions and activities; especially in domains of skilled performance.

Recognizing these dependencies the importance of defaults in action selection has to be focused. Using defaults allows automatic behavior by reducing effort necessary for information acquisition and weighing different courses of action (Johnson & Goldstein, 2003). According to O'Hare (2003), "it will be easier to continue with an existing course of action than to change to a new one" (p. 223). So pilots will sometimes tend to stick to unsuitable skill- or rule-based behavior, where analytical knowledge-based strategies would be appropriate (O'Hare, 2003).

Based on these considerations of the interplay of system 1 and system 2 there is one central conjecture on pilot behavior: Variability is to be expected in the influence of system 1 and system 2 on decision making. This variability leads to different degrees of endorsement of less effortful behavioral or decision strategies. We suppose two possible consequences of these strategies: Variability in investing effort in data acquisition behavior and variability in sticking to default decision options. We suppose that these behavioral tendencies have clear influences on the decision to go around.

Hypotheses

Research Hypothesis 1: Pilots with a high level of expertise will come to 'better' decisions based upon good airmanship.

The consequences of potential effort reducing strategies described above might become manifest in different gaze strategies for pilots that finally take the decision to go around (PGA) versus those pilots that would presumably take the decision to land in spite of strong tailwind and a cue from the PM (PnGA). The difference between these two groups might stem from different information acquisition strategies or from different use of default decision options. According to this demarcation, two different, mutually exclusive gaze behaviors would result as a consequence:

Research Hypothesis 2: Pilots not intending to go around (PnGA) perceive relevant information too late or not at all. That is expressed by the following gaze profile: First wind check is later for PnGA than for PGA. There is no difference in final wind check between PnGA and PGA. There is a difference between first and final wind check between PnGA and PGA.

Research Hypothesis 3: Pilots intending to land (PnGA) stick to a default option; up to the point of deciding to choose the default of landing, information acquisition does not differ from PGA. That is expressed by the following gaze profile: First wind check is not differing between PnGA and PGA. There is a difference in final wind check between PnGA and PGA. There is a difference between first and final wind check between PnGA and PGA (Figure 2).

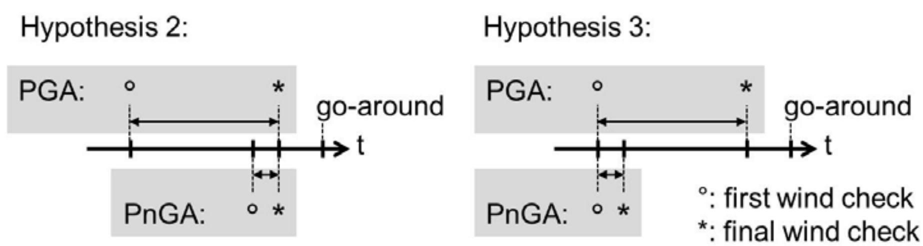


Figure 2. Time charts for wind checks concerning hypotheses 2 and 3.

Method

Participants

This study has been undertaken in cooperation with a major European airline. Pilots with different levels of practice and training were scheduled for a flight simulator experiment by their operations department; i. e. participation was not voluntary. Twenty-six long-haul captains (CPTs) flying Airbus A330/340 types participated the experiment in a full flight simulator (JAR-STD 1A Level D) with A340-600 configuration and twenty-seven first officers (FOs) scheduled on the A 320 short-haul fleet participated in an equivalent A320-200 full flight simulator. The CPTs had a lower level of practice and training, because of their flight school attendance was more way back and they had only few long-haul operations per month (mean value of own performed landings in the 30 days prior to the experiment = 3.4). As contrasting group, younger FOs coming recently from flight school face a high number of short-haul operations per month (mean value of own performed landings in the 30 days prior to the experiment = 16.6). Flight experience (total flight hours) is diametrically opposed to the level of practice and training. All participants had the role of the (PF). A confederate PM complemented the aircrew and was instructed to play a rather passive roll but to avoid errors. All pilots were scheduled on the airline's corresponding fleet, held an appropriate license (ATPL) and were asked to prepare themselves in a same way as for a real flight, to wear uniforms and to bring their own computer for the electronic flight bag system (Haslbeck et al., 2012).

Scenario

All participants were briefed on an uneventful flight from the east to Munich Airport in the early morning hours. The PF came back from his last rest about 25 minutes prior to the landing to perform the approach and landing. In the first phase of the approach, using the autopilot, foreign air traffic control (ATC) communication ('party line') between other approaching aircraft and the airport could be heard. Pilots' tasks were to plan, to monitor and to communicate. When approaching the instrument landing system, it was the PF's decision when to change from autopilot to manual control. To provoke a hazardous situation, at 1.000 ft. above ground level (AGL), a gentle wind turned into an illegitimate strong tailwind (16 knots) by a scripted event. The wind information given by ATC was constantly good over the whole scenario. For pilots, this information given by ATC is binding. Only the non-binding wind indicator located at the pilot's navigation display has shown the real wind strength and direction. Such a hazardous situation can occur when the wind turns, because the wind information given by ATC is averaged over several minutes. So the situation was inexplicit and uncertain for the participants to make the trade-off between a fuel-saving and economic landing with a noticeable higher risk or the abort of the approach for a safe second try (Haslbeck et al., 2012). The chance to go around was given to all participants until 70 ft. AGL. At this height, the PM was instructed to callout 'go-around' and abort the approach due to strong tailwind.

Measurement

Behavioral data were recorded by three complementary methods. All participants were equipped with a head-mounted eye-tracking system (DIKABLIS) to measure their visual behavior (Haslbeck, Schubert, Gontar, & Bengler; 2012); pilots' control inputs were recorded by the flight simulator's data recorder; finally, video and audio data in the cockpit were recorded. From these different sources, a comprehensive image of pilots' decision making can be drawn.

Results

Accomplishment of the go-around according to the different types described in Table 1 was analyzed; results are shown in Table 2. The distinction between type 1 and 2 is based upon the gaze data. It was not possible to distinguish between type 3 and 4, because these pilots could not be asked whether they were aware of the wind in that explicit situation. Pilots were asked about this in the debriefing session after the experiment, but not all could clearly recall this situation and thus type 3 and 4 are considered as one joint type here.

Table 2.

Accomplishment of the go-around.

	Type 1	Type 2	Type 3/4	Total
CPTs	10	4	12	26
FOs	14	11	2	27

In a statistical comparison between CPTs and FOs, significant differences between both groups were found using the chi-squared test: $\chi^2(2)=11.02$; $p = .004$. Because two cells contained less than five cases, Fisher Exact Probability Test was additionally used. This test also shows significant results ($p = .004$).

Comparison of gaze behavior of pilots who took the decision to go around (PGA) and pilots intending to land (PnGA) led to distinct results. Statistical analysis was carried out using Mann-Whitney-Wilcoxon test. Effect size is expressed in units of a bivariate correlation coefficient *r*, as suggested by Rosenthal and Rosnow (2007). The time of first wind check (t_1) was markedly earlier for PGA than for PnGA: $U = 89$; $z = 2.11$; $p = .035$; $r = .399$, indicating a medium to large effect. There was no significant difference in regard of the time of last wind check (t_2): $U = 137.5$; $z = 0.54$; $p = .597$; $r = .102$. The difference between the time of first and last wind check showed a considerable difference: $U = 49$; $z = 3.41$; $p < .001$; $r = .645$, indicating a very large effect.

Discussion

The results of this study show that long-haul captains with a lower level of practice and training but a high level of operational experience show significantly more willingness to land in a risky situation with strong tailwind than short-haul first officers do. Thus Hypothesis 1 is supported/not supported by the data. When thinking about this behavior, the question arises, how a 'better' decision can be characterized. Two different statements were given by PnGAs corresponding to type 3 and 4:

Type 3: The pilot was aware of the tailwind, but decided to land. In the concrete situation of Munich Airport (MUC) a tailwind landing may be an acceptable risk for someone. Both runways have a length of 4.000m each, which offer a certain safety margin (for comparison longest runway 15R at Boston 3.073m). In addition, some pilots are aware of the fact, that performing a missed approach is also a challenge after several hours of flight duty and so they tend to avoid the go-around.

Type 4: The pilot was not aware of the tailwind and thus the risk of this situation. Some reasons can be fatigue, high workload or a complacent behavior towards the wind situation because of safe wind information given by ATC. This case means insufficient airmanship and yields a higher risk in aviation, independently from the location of the airport.

Hypothesis 1 can neither be accepted nor rejected entirely. Under the assumption, the 'better' decision was to go around, the FOs with a higher level of practice and training but less experience more often made the 'better' decision to go around. Under the assumption that pilots who feel uncomfortable with a go-around under these circumstances (type 3), the results are not that clear. In this case, more CPTs came to a 'better' decision from their point of view. One limitation of this analysis is the fact, that type 3 and 4 couldn't be clearly distinguished after the experiment.

Results of gaze behavior analysis support the general assumption that there is a difference in effort investment between pilots who took the decision to go around (PGA) and pilots intending to land (PnGA). Based on dual-process theory of cognition two distinct patterns of gaze behavior were derived. According to *Research Hypothesis 2* effort-preserving behavior leads PnGA to delayed information acquisition and so to different first but comparable final wind-checking times. According to *Research Hypthesis 3* reliance on a default option makes PnGA to comparable first but different final wind-checking times. Results provide evidence for the first of these two hypotheses. PnGA perceive relevant information markedly later than PGA, whereas there is no difference in time of the last check. Although not applicable to differentiate between the two explanatory approaches, the difference in time of first and last wind check between PGA and PnGA serves as strong support for the general assumption of effort preserving behavior; be it by way of strategic decision behavior in form of information acquisition or in form of sticking to a default alternative. There seems to be variability in the influence of system 1 and system 2 on decision making. This variability leads to different degrees of endorsement of less effortful behavioral or decision strategies, in turn leading to different information acquisition strategies for PGA and PnGA.

These results suggest two leverage points for supporting decision quality. Effort investment could be reduced by shifting the balance between system 1 and system 2 processes. Information acquisition is cognitively effortful in part because it is not fully automatized. Intensive gaze training procedures could improve theses information acquisition skills. Another approach is to use findings of the fields of human-computer interaction and human factors, to reduce acquisition effort by designing displays appropriately; e. g. by reducing the gaze angle necessary or making relevant information visually more distinct.

Interestingly only a small number of pilots (26 %) would have landed without the PM being instructed to trigger the go-around in any case. Instructor pilots normally report a higher tendency to go-around when being in the flight simulator in comparison to real flights. Pilots can show safety awareness in the flight simulator by frequently performing go-arounds without really making the trade-off between safe flight operation and economic constraints, while in reality the tendency to go around seems lower.

Acknowledgements

This work was funded by the German Federal Ministry of Economics and Technology via the Project Management Agency for Aeronautics Research within the Federal Aeronautical Research Program (LuFo IV-2). The authors acknowledge their thanks to Patrick Gontar and Ekkehart Schubert for supporting this study and their contribution to the project.

References

Boeing Commercial Airplanes. (2012). *Statistical Summary of Commercial Jets Airplane Accidents: Worldwide Operations 1959 - 2011*. Seattle. Retrieved from http://www.boeing.com/news/techissues/pdf/statsum.pdf

DaMaria, C. (2006). *Understanding Airmanship*. Retrieved from http://www.aviationchannel.com/article/article.php?id=5

Endsley, M. R. (2006). Situation Awareness. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 528–546). Hoboken, NJ, USA: John Wiley & Sons, Inc.

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. doi:10.1146/annurev.psych.59.103006.093629

Federal Aviation Administration. (2004). *Airplane Flying Handbook* (No. FAA-H-8083-3A). Retrieved from http://www.faa.gov/library/manuals/aircraft/airplane_handbook/

Haslbeck, A., Schubert, E., Gontar, P., & Bengler, K. (2012). The relationship between pilots' manual flying skills and their visual behavior: a flight simulator study using eye tracking. In S. Laundry, G. Salvendy, & W. Karwowski (Eds.): *Advances in Human Factors and Ergonomics, Advances in Human Aspects of Aviation* (pp. 561–568). Boca Raton: CRC Press.

Haslbeck, A., Schubert, E., Onnasch, L., Hüttig, G., Bubb, H., & Bengler, K. (2012). Manual flying skills under the influence of performance shaping factors. *Work: A Journal of Prevention, Assessment and Rehabilitation*, *41*(Supplement 1/2012), 178–183. Retrieved from http://dx.doi.org/10.3233/WOR-2012-0153-178

International Air Transport Association. (2011). *Safety Report 2010*. Issued April 2011. Montréal.

Jenkins, D. P., Stanton, N. A., Walker, G. H., Salmon, P. M., & Young, M. S. (2008). Applying cognitive work analysis to the design of rapidly reconfigurable interfaces in complex networks. *Theoretical Issues in Ergonomics Science*, *9*(4), 273–295. doi:10.1080/14639220701561833

Johnson, E. J., & Goldstein, D. (2003). Do Defaults Save Lives? *Science*, *302*(5649), 1338–1339. doi:10.1126/science.1091721

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review*, *93*(5), 1449–1475. doi:10.1257/000282803322655392

Kahneman, D. (2011). *Thinking, fast and slow*. London: Allen Lane.

Kahneman, D., Frederik, & S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases. The psychology of intuitive judgment* (pp. 49–81). Cambridge, U.K, New York: Cambridge University Press.

O'Hare, & D. (2003). Aeronautical decision making: metaphors, models, and methods. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practice of aviation psychology* (pp. 201–237). Mahwah, N.J: Lawrence Erlbaum.

Orasanu, J. (1995). Situation Awareness: Its role in Flight Crew Decision Making. 8th International Symposium on Aviation Psychology.

Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-13*(3), 257–266.

Rosenthal, R., & Rosnow, R. L. (2007). *Essentials of behavioral research: Methods and data analysis* (3rd). Boston: McGraw-Hill.

MODELING HUMAN AND ANIMAL COLLISION AVOIDANCE STRATEGIES

Ioana Koglbauer, Reinhard Braunstingl, and Thomas Haberkorn
Institute of Mechanics, University of Technology, Graz, Austria

In this paper we propose technical solutions for overcoming pilots' limitations in handling collision situations in visual flight. An analysis of pilots' requirements for future development of such systems shows that a surprisingly large proportion of pilots would prefer automated systems which will allow decision and performance of the avoidance maneuvers, and recapturing of the initial route by the autopilot. For building up a model of collision avoidance we reviewed previous findings on collision avoidance strategies of pilots, air traffic controllers and pedestrians. Additionally, we reconsidered studies on collision avoidance strategies of birds and insects. Finally, we discuss new challenges for future cockpit technology and flight training methods to improve collision avoidance management and safety of VFR pilots.

Making use of powered vehicles humans navigate in spacetime zones to which they were not predestined by their natural abilities. In their natural eco-systems animals and humans are adapted to protect their personal space from intrusion. However, as accident and incident related to the use of powered vehicles show, humans are vulnerable to collisions within the artificially created spacetime despite existent technology, procedures and training.

The rate of midair collisions involving pilots flying according to visual flight rules (VFR) remained constant despite an overall decline of General Aviation accidents in the past decades (AOPA, 2009). An analysis of midair collisions in VFR (Taneja & Wiegmann, 2001) showed that all midair collisions occurred under visual meteorological conditions, 35.4 % involved a head-on configuration and many resulted from an overtaking maneuver. In 85% of cases the reported cause of accident was inadequate look out by the pilot(s).

There are many factors which make detection of interpretation of conflict situations difficult, despite good visual meteorological conditions. However, technological solutions such as the new collision warning systems for VFR (e.g. FLARM and ADS-B based systems) which display traffic and warnings are expected to improve flight safety. We implemented such a system in flight simulator (www.flightsimulation.tugraz.at) at Graz University of Technology and evaluated the benefits and limitations of new collision warning systems for VFR pilots.

In the following we present briefly the main results of our study and compare findings from the literature which evaluated collision avoidance preferences of pilots, air traffic controllers and pedestrians. We also mention avoidance strategies of insects and birds. These are excellent flyers, naturally adapted to live in flocks or swarms, and manage to fly safely without recurrent proficiency checks, instruments, or warnings. Finally we draw out implications for modeling human and animal collision avoidance behavior.

On collision avoidance tactics in humans and animals

Study of pilots' initial collision avoidance responses in visual flight

In a simulator study we evaluated the initial responses of 18 VFR pilots to single and multiple traffic approaching at the same altitude (see also Haberkorn, Koglbauer, Braunstingl & Prehofer, 2013). Traffic was displayed on the visual system of the simulator and on a moving map display which generated a warning 24 seconds before the closest point of approach. Pilots were flying at 6000 ft above ground level.

Pilots' responses for two conflict geometries are illustrated in Figures 1 and 2. A short remark for the interpretation: according to aviation regulations the correct evasive maneuver in case of frontal conflicts is a turn to the right, meaning in our case a positive change of heading. Approaching aircraft (from the left, 290 degree) did not respect the priority rules being non-cooperative and forcing the pilots to induce necessary avoidance maneuvers. As illustrated in Figure 1, where the oncoming aircraft was approaching head-on from 10°, preferences for descends combined with evasive turns to either left or right can be noticed. In the multiple traffic condition (Figure 2) pilots' preference for descents and a right turns can be observed.

Traffic load influenced significantly pilots' reactions. The amplitude of pitch changes and the changes of vertical speed regardless of direction were higher for multiple conflicts than for single conflicts. The magnitude of heading changes decreased from the first to the second and third trail. In conditions of multiple traffic reaction times to traffic warning were longer, mental workload increased, whereas spare mental capacity and anticipation decreased.

Thomas and Wickens (2005) also found a preference for vertical evasive maneuvers in level conflicts with approach angles of less than 120°. Climbs were most preferred (45%) followed by level flight (31%) and descents (25%). Only in 55% of cases pilots avoided to the right. Actually, a large amount of previous studies with pilots confirms a preference for vertical maneuvers (Abbott, Moen, Person, Keyser, Yenni & Garren, 1980; Merwin & Wickens, 1996; O'Brien & Wickens, 1997; Wickens & Morphew, 1997; Gempler & Wickens, 1998; Wickens & Helleberg, 1999; Helleberg, Wickens & Xu, 2000, Alexander & Wickens, 2001). Palmer (1983) reported a preference for vertical versus lateral only under time pressure. However, results are not conclusive. There are also studies showing preferences for lateral collision avoidance maneuvers (Palmer, 1983; Ellis, McGreevy & Hitchcock, 1987). Only few studies showed preferences for changing airspeed maneuvers (Chappel & Palmer, 1983).

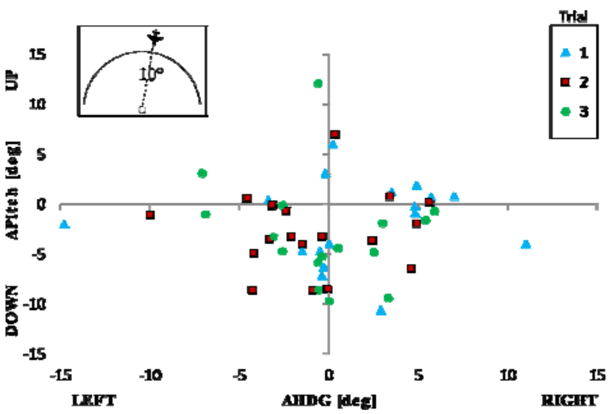


Figure 1. Pilot initial responses to an aircraft approaching from 10 degree at the same altitude. Changes of pitch and heading, 5 seconds after reaction begin.

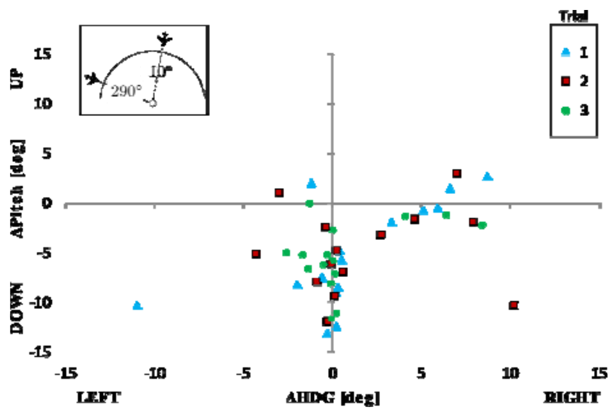


Figure 2. Pilot initial responses to aircraft approaching from 10 and 290 degree at the same altitude. Changes of pitch and heading, 5 seconds after reaction begin.

Collision avoidance preferences of air traffic controllers and pedestrians

Studies with air traffic controllers showed similar results. Rantanen, Yang and Yin (2006) found a preference of student controllers for climb (39.9%) and descend (35.7), whereas speed adjustment was used in 19.4% of cases (increase 12.8% and decrease 6.5%). Heading changes were used only 5% with no differences between left and right.

The above mentioned studies address human interactions with artificial systems and spacetime dynamics which exceed their natural environment. For getting a better picture of "natural" human collision avoidance preferences we are also investigated empirical findings with pedestrians. Interestingly, studies of pedestrians avoidance strategies show preferences for speed changes (Blue, Embrechts & Adler, 1997), maintenance of current direction (Antonini, Bierlaire, and Weber, 2006) or minimization of the angular displacement (Turner and Penn, 2002) and smooth non-linear trajectories (Bierlaire et al. 2003). If a head-on collision is imminent both pedestrians tend to make a side-step (Helbing, Molnar, Farkas, and Bolay, 2001). Some findings show that pedestrians tend to pass on the right in head-on conflicts (Goffman, 1971), but other reported that pedestrians passed on both left and right with equal probability (Daamen and Hoogendoorn, 2003).

The tendency to minimize heading changes after more repetitions of the same avoidance maneuver was found in both our study with pilots and studies with pedestrians (Antonini, Bierlaire, and Weber, 2006; Turner and Penn, 2002). These findings suggest a cost-benefit evaluation in choosing the avoidance maneuver. Together with the general preference for vertical maneuvers these findings indicate a tendency to protect the initial route or an economic bias in choosing the avoidance tactic.

Collision avoidance strategies of birds and insects

For a better insight in evolutionary-proven strategies we searched for information on animal collision avoidance behavior. We know that birds and insects are well adapted to maintain separation from their flock neighbors and obstacles or to escape predators. Although their brain is simpler than the human brain, their 3D collision detection and avoidance mechanisms seem to be more effective (Miller, Ngo & van Swinderen, 2012). This was the reason for development of new bio-inspired collision detection systems (Stafford, Santer & Rind, 2007a; 2007b).

Interestingly, we found that studies with locusts also show a preference for vertical evasions. For escaping predators and for collision avoidance with another locust during flight locusts perform a gliding dive, ceasing to beat their wings and holding all four wings raised with above the hind wings held slightly swept back (Santer, Simmons & Rind, 2005). During diving locusts decelerate from 4 m/s to 2.5 m/s and lose about 3.5 cm height. For collisions at relative approach speed slower than 3 m/s the dive is preceded and accompanied by other steering movements (Simmons, Rind & Santer, 2010).

The pea aphid (*Acyrthosiphon pisum*) also uses dropping as anti-predator tactic (Lawrence, Fraser & Roitberg, 1990). Research shows that pea aphids are less likely to drop and avoid when they have high quality feeding hosts or when they are in hot and dry conditions and the risk of desiccation is higher. Thus, their tactic can be interpreted in an economic cost-benefit framework. Similarly, the delay of first escape responses of redshank (*Tringa totanus*) flocks was found to be determined by economic reasons. Delay was increasing for low-risk stimuli and decreasing for high-risk stimuli such as hawks (Quinn & Cresswell, 2005).

Another interesting aspect of animal behavior is the spatial organization of flocks. Lukeman, Li and Edelstein-Keshet (2010) observed native undomesticated flocking surf scoters in groups of up to 200 individuals in their natural setting, on water surface. In the horizontal plane scoters changed their speed and deviated sideways for maintaining separation in the flock. They made frequent +/-180 degree head turns to watch their neighbors. The spatial behavior in the flock could be described in terms of short-range repulsion for maintaining separation, intermediate-range alignment and longer-range attraction of the individuals to the flock. The individual position, velocity and trajectory created spatial and angular neighbor distribution plots, showing a concentric structure in positioning, a preference for neighbors directly in the front, and strong alignment with neighbors on each side. Furthermore, scoters were structuring their space to form empty avoidance zones which they use to escape encroaching gulls which attack them to rob their mussels.

We see that flying insects and birds seem to use primarily the vertical plane to escape. Pedestrians and surfing birds prefer speed changes and horizontal escape maneuvers with minimal deviance from the route. Expected costs and benefits seem to influence the deviance from the route and the delay of the escape maneuver.

Considerations for modeling collision avoidance strategies

In our study requirements for future development of collision avoidance systems in VFR were specified with feedback from pilots (see also Haberkorn et al., 2013). Automated aids for detecting traffic conflicts were required by 94.4% of the pilots, whereas automatic generation of an avoidance route was required by 61%. With regard to automatic performance of the avoidance maneuver 58.8% of pilots would prefer the option to abort and 44.4% would prefer the option to approve an automatic evasive maneuver. With regard to automatic recapturing of the initial route after performing the avoidance maneuver 61.1% would prefer the option to abort and 72.2% would prefer the option to approve an automatic maneuver.

From the kinematic point of view collision situations are precisely described in terms of time and distance to collision of two or more vehicles approaching with specific speeds, and optimal routes for collision avoidance can be calculated by algorithms. However, in nature detection and avoidance of collisions is based on intuitive mechanisms. We consider that automatic collision avoidance systems will be accepted only if they do not contradict the intuitive judgment of humans. Thus we attempt to apply basic knowledge about intuitive strategies applied by humans and animals for future development of collision avoidance systems. If pilots will either abort or approve an automatic collision avoidance maneuver, they should be able to evaluate conflict situations and automatically calculated routes in a timely manner.

TCAS reports from commercial aviation show that pilots do not always comply with automatically generated resolution advisories (RA). Coso, Fleming and Pritchett (2011) found that in 8% of 251 cases pilots have not accepted the RA. Mentioned reasons were the contradiction of pilots' attempt to keep the intruder aircraft in sight, the involvement of a third aircraft, or pilots' impression that the RA would direct them into traffic. In other cases pilots added a horizontal component to the RA, which are solely vertical. As Coso et al. (2011) showed, pilots judge more contextual information than TCAS and these judgments influence their compliance with RAs.

Implications for modeling automatic collision avoidance strategies

We propose to include both contextual information and collision avoidance preferences in modeling collision avoidance strategies of future automatic systems.

Contextual information:

- Prioritization in conditions of multiple traffic
- Consideration of non-sensed flying objects and fixed obstacles, terrain, airspace restriction data
- Rules and regulations (e.g. priority according to type of air vehicle, conflict geometry, airport procedures)
- Safety envelope of the aircraft (e. g. airspeed, altitude, pitch, roll, yaw)
- Autopilot features
- Data-link: communication aircraft-aircraft, aircraft - air traffic control

Collision avoidance preferences:

- Display indication: multimodal display (visual and acoustic, to avoid increasing head-down time)
- Timing – for evaluation of both the conflict situation and the proposed automatic avoidance maneuver
- Automation approval and /or override options
- Route protection (minimal heading changes or attraction to the next leg of the route, speed changes)
- Use of a vertical component
- Smooth, non-linear trajectory

Trajectory building

The goal of any trajectory planning algorithm is to determine future history of control inputs on the own aircraft under influence of minimizing a cost function (Stengel, 1994). This mechanism should be robust against disturbances such as partial signal loss and wind turbulence.

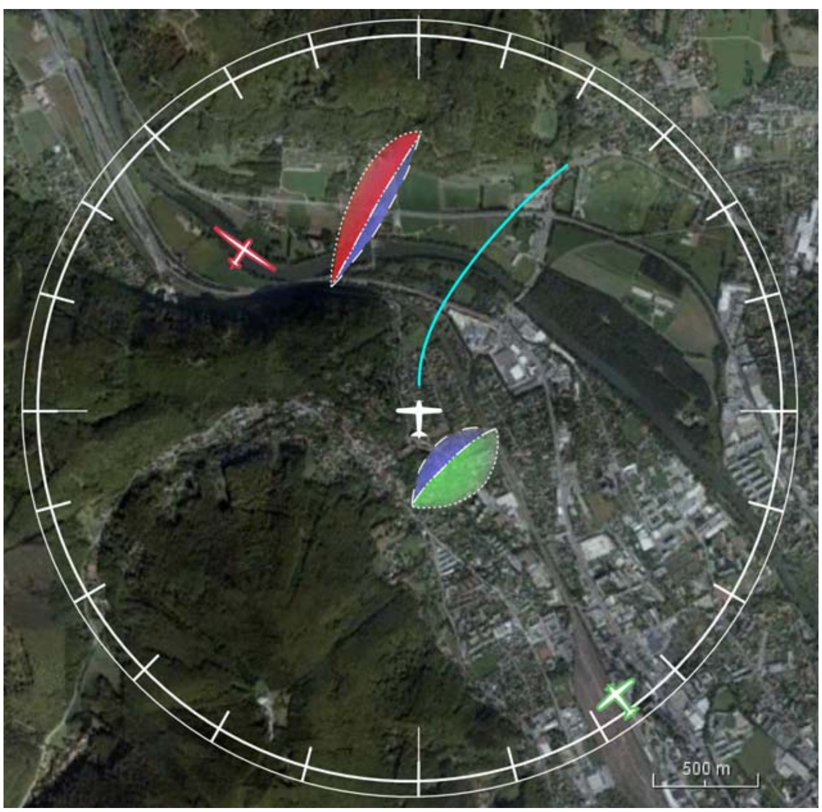


Figure 3. Possible horizontal trajectory to avoid conflict with the glider approaching from the left and a with a powered aircraft from behind by using the volume of intersecting sets as a cost function.

One way to calculate the cost function is to determine volumetric protected zones around each participating aircraft, with shapes and sizes dependend on specific aircraft capability. The cost function is calculated using the volume of intersecting set of protected zones. From all possible control inputs to the own aircraft, candidate-trajectories are selected. These are optimal trajectories in the sense of minimizing the volume of intersecting sets. Trajectories have to be further compared according to context criteria stated above (e. g. rules and regulations, route protection). Furthermore, the energy efficiency of remaining candidate trajectories has to be compared.

In addition the system may use a pilot setup for the pre-selected mode of operation (e.g. horizontal evasion at constant velocity, vertical evasion, evasion by speed changes). Filtering of remaining alternatives could then be performed according to the pre-selected mode. Finally, a trajectory will be chosen and displayed by the system (Figure 3). Pilots will be able to control automation using the approval and override options.

Final Considerations

Although automated collision avoidance in VFR is possible, we expect to encounter certain limitations for future systems. The main limitation originates in the characteristics of airspace destined for VFR flight. Models of obstacles, terrain and airspaces should be considered, as well as sensors for detecting all flying objects, not only those equipped with FLARM, ADS-B or transponder.

Furthermore, we need more research to evaluate pilots' interaction with automatic collision avoidance systems. With respect to human bias in selecting collision avoidance options (e.g. vertical, lateral or airspeed maneuvers) we know little about pilots' acceptance, monitoring and trust in such alternatives when they are proposed by automation.

Additionally, hands-on simulator training for handling automatic collision avoidance systems will be necessary. Training scenarios should include also scenarios with imperfect automation (e.g. flying objects not recognized by the system etc.). For improving safety and efficiency automatic collision avoidance systems should be used and controlled by pilots who are familiar with benefits and limitations of automation.

References

Abbott, T. S., Moen, G. C., Person, L. H., Keyser, G. L., Yenni, K. R., & Garren, J. F. (1980). *Early flight test experience with cockpit displayed traffic information.* NASA Technical Memorandum 80221 /AVRADCOM Technical Report 80-B-2, Hampton, VA, NASA Langley.

Alexander, A., & Wickens, C. D. (2001). *Cockpit display of traffic information: the effects of traffic load, dimensionality, and vertical profile orientation.* University of Illinois Institute of Aviation Technical Report (ARL-01-17/NASA-01-08). Savoy, IL: Aviation Res Lab.

Antonini, G., Bierlaire, M., & Weber, M. (2006). Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological*, 40(8), 667-687.

AOPA Air Safety Foundation (2009) Collision avoidance. Strategies and Tactics. *Safety Advisor. Operations and Proficiency*, 4. [Downloaded on 2nd October 2011 at www.asf.org].

Bierlaire, M., Antonini, G., & Weber, M. (2003). Behavioral dynamics for pedestrians. In K. Axhausen (Ed.), *Moving through nets: The physical and social dimensions of travel*, Elsevier, 1-18.

Blue, V. J., Embrechts, M. J., & Adler, J. L. (1997). Cellular automata modeling of pedestrian movements. *IEEE International Conference on Systems, Man and Cybernetics*, 2320-2323.

Chappel, S. L., & Palmer, E. A. (1983). Primary separation between three aircraft using traffic displays. *Proceedings of the Human Factors Society 27th Annual Meeting*, Santa Monica, CA, 767-771.

Coso A. E., Fleming E. S., & Pritchett A. R. (2011). Characterizing pilots' interactions with the aircraft collision avoidance system. *Proceedings of the 16th International Symposium on Aviation Psychology*, Dayton, OH, 493-498.

Daamen, W., & Hoogendoorn, S. P. (2003). Experimental research of pedestrian walking behavior. *Transportation Research Board Annual Meeting*, National Academy Press, 1-16.

Ellis, S. R., McGreevy, M. W., & Hitchcock, R. J. (1987). Perspective traffic display format and airline pilot traffic avoidance. *Human Factors*, 29, 371-382.

Gempler, K. S., & Wickens, C. D. (1998). *Display of predictor reliability on a cockpit display of traffic information.* University of Illinois Institute of Aviation Final Technical Report (ARL-98-6/ROCKWELL-98-1). Savoy, IL: Aviation Res Lab.

Goffman, E. (1971). *Relations in public: microstudies of the public order.* New York, Basic Books.

Haberkorn, T., Koglbauer, I., Braunstingl, R., & Prehofer, B. (2013). Pilots' requirements for future collision avoidance systems in visual flight: A human-centered approach (submitted).
Helbing, D., Molnar, P., Farkas, I. J., & Bolay, K. (2001). Self-organizing pedestrian movement. *Environment and Planning B: Planning and Design*, 28, 361-383.
Helleberg, J., Wickens, C. D., & Xu, X. (2000). *Pilot maneuver choice and safety in a simulated free flight scenario*. University of Illinois Institute of Aviation Technical Report ARL-00-1/FAA-00-1. Savoy, IL: Aviation Res Lab.
Lawrence, M. D., Fraser, A. H. G., & Roitberg, B. D. (1990). The economics of escape behaviour in the pea aphid, Acyrthosiphon pisum. *Oecologia*, 83(4), 473-478.
Lukeman, R., Li, Y.-X., & Edelstein-Keshet, L. (2010). Inferring individual rules from collective behavior. *Proceedings of the National Academy of Science*, 107(28), 12576-12580.
Merwin, D. H., & Wickens, C. D. (1996). *Evaluation of perspective and coplanar cockpit displays of traffic information to support hazard awareness in free flight.* University of Illinois Institute of Aviation Technical Report (ARL-96-5/NASA-96-1). Savoy, IL: Aviation Res Lab.
Miller, S. M., Ngo, T. T., & van Swinderen, B. (2012). Attentional switching in humans and flies: rivalry in large and miniature brains. *Frontiers in Human Neuroscience*, 5, 188, 1-17.
O'Brien, J. V., & Wickens, C. D. (1997). Free flight cockpit displays of traffic and weather: effects of dimensionality and data base integration. *Proceedings of the Human Factors and Ergonomics Society, 41st Annual Meeting*, Santa Monica, CA, 18-22.
Palmer, E. A. (1983). Conflict resolution maneuvers during near miss encounters with cockpit traffic displays. *Proceedings of the Human Factors Society 27th Annual Meeting*, Santa Monica, CA, 757-761.
Quinn, J. L., & Cresswell, W. (2005). Escape response delays in wintering redshank, Tringa totanus, flocks: perceptual limits and economic decisions. Animal Behaviour, 69, 1285-1292.
Rantanen, E. M., Yang, J., & Yin, S. (2006). Comparison of pilot's and controllers' conflict resolution maneuver preferences. *Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA.
Santer, R., Simmons, P. J., & Rind, F. C. (2005). Gliding behaviour elicited by lateral looming stimuli in flying locusts. *J Comp Physiol A Sens Neural Behav Physiol*, 191, 61-73.
Simmons, P. J., Rind, F. C., & Santer, R. (2010). Escapes with and without preparation: the neuroethology of visual startle in locusts. *Journal of Insect Physiology*, 56, 876-883.
Stafford, R., Santer, R. D., & Rind, F. C. (2007a). The role of behavioural ecology in the design of bio-inspired technology. *Animal Behaviour*, 74, 1813-1819.
Stafford, R., Santer, R. D., & Rind, F. C. (2007b). A bio-inspired visual collision detection mechanism for cars: combining insect inspired neurons to create a robust system. *BioSystems*, 87, 164-171.
Stengel, R. F. (1994). *Optimal control and estimation.* Dover Publ. Inc, New York.
Taneja, N., & Wiegmann, D. A. (2001). Analysis of mid-air collisions in civil aviation. *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA.
Thomas, L. C. and Wickens, C. D. (2005). Display dimensionality and conflict geometry effects on maneuver preferences for resolving in-flight conflicts. *Proceedings of The Human Factors and Ergonomics Society Annual Meeting*, 1, 40-44.
Turner, A., & Penn, A. (2002). Encoding natural movement as an agent-based system: An investigation into human pedestrian behaviour in the built environment. *Environment and Planning B: Planning and Design*, 29, 473-490.
Wickens, C. D., & Helleberg, J. (1999). *Interactive perspective displays for airborne hazard awareness.* University of Illinois Institute of Aviation Final Technical Report (ARL-99-1/ROCKWELL-99-1). Savoy, IL: Aviation Res Lab.
Wickens, C. D., & Morphew, E. (1997). *Predictive features of a cockpit traffic display: A workload assessment.* University of Illinois Institute of Aviation Technical Report (ARL-97-6/NASA-97-3). Savoy, IL: Aviation Res Lab.

Acknowledgments

We gratefully acknowledge the pilots who participated in our study and Boris Prehofer for his support in setting-up the flight simulator for generating multiple traffic scenarios.

Aviation safety evaluation by wavelet kernel-based support vector machine

Chang ming , You Xuqun
*School of Psychology, Shaanxi Normal University,Xian 710062,China,
cchangmingg@163.com

Abstract

In order to obtain the excellent evaluation effects, the wavelet kernel function is used as the kernel function of support vector machine,and the model is defined as wavelet kernel-based support vector machine.Thus, wavelet kernel-based support vector machine is applied to aviation safety evaluation. The two dimensional input vector of the training samples is employed to construct the training samples.The traditional radial basis function kernel-based support vector machine is used to compare with the wavelet kernel-based support vector machine.The testing results show that the evaluation error of the wavelet kernel-based support vector machine belongs to the range from 0.015 to 0.04,and the evaluation error of the traditional radial basis function kernel-based support vector machine belongs to the range from 0.02 to 0.07.Then,we can conclude that aviation accidents evaluation accuracy of wavelet kernel-based support vector machine is higher than those of traditional radial basis function kernel-based support vector machine.

Keywords: *wavelet kernel; evaluation;aviation safety;support vector machine*

1.Introduction

Recently, artificial neural networks have been used less in engineering field. Instead of artificial neural networks, support vector machine has been applied for various domains, due to its excellent generalization ability[1-6].Support vector machine proposed by Vapnik is a novel machine learning algorithm based on statistical learning theory[7,8]. Originally,support vector machine was developed to solve the classification problems. Recently, the method has been extended to solve the regression and evaluation problems.The choice of kernel function of support vector machine has a certain influence on the evaluation effects.In order to obtain the excellent evaluation effects, the wavelet kernel function is used as the kernel function of support vector machine,and the model is defined as wavelet kernel-based support vector machine.Thus, wavelet kernel-based support vector machine is applied to aviation safety evaluation. Aviation accidents from 1995 to 2003 in China are employed as our experimental data to test and analysis for aviation accidents[9] evaluation by using the proposed method.The two dimensional input vector of the training samples is employed to construct the training samples.The traditional radial basis function kernel-based support vector machine is used to compare with the wavelet kernel-based support vector machine.The testing results show that the evaluation error of the wavelet kernel-based support vector machine belongs to the range from 0.015 to 0.04,and the evaluation error of the traditional radial basis function kernel-based support vector machine belongs to the range from 0.02 to 0.07.Then,we can conclude that aviation accidents evaluation accuracy of wavelet kernel-based support vector machine is higher than those of traditional radial basis function kernel-based support vector machine.

2.Wavelet Kernel-based Support Vector Machine

Support vector machine proposed by Vapnik is a novel machine learning algorithm based on statistical learning theory.Recently, the method has been extended to solve the regression and evaluation problems[10,11].

Given a set of the training samples $\{(x_i, y_i)\}_{i=1}^{n}$, where $x_i \subset R^m$ denotes the input vector and $y_i \subset R$ denote the corresponding output. The evaluation function of support vector machine employs the following formula:

$$y(x) = \omega \cdot \phi(x) + b \tag{1}$$

where ω denotes the weight vector and b denotes the bias.

In order to gain the values of the weight vector ω and the bias b, two positive slack variables ξ, ξ^* are introduced,and infeasible constraint of the optimization problem is given as follows:

Minimize

$$\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) \tag{2}$$

Subject to

$$\begin{cases} y_i - \langle \omega, \phi(x_i) \rangle - b \le \varepsilon + \xi_i & \xi_i \ge 0 \\ \langle \omega, \phi(x_i) \rangle + b - y_i \le \varepsilon + \xi_i^* & \xi_i^* \ge 0 \end{cases}$$

where C denotes the penalty.

Then, the Lagrangian multipliers a_i, a_i^* are introduced to obtain the constrained optimization problem:

Maximize

$$\sum_{i=1}^{n} y_i\left(a_i - a_i^*\right) - \varepsilon\sum_{i=1}^{n}\left(a_i + a_i^*\right) - \frac{1}{2}\sum_{i,j=1}^{n}\left(a_i - a_i^*\right)\left(a_j - a_j^*\right)K(x_i, x_j) \tag{3}$$

Subject to

$$\begin{cases} \sum_{i=1}^{n}(a_i - a_i^*) = 0 \\ 0 \le a, a^* \le C \end{cases}$$

where $K(x_i, x_j)$ ($K(x_i, x_j) = \phi(x_i)\phi(x_j)$) denotes the kernel function.

Finally, the evaluation function of support vector machine is given as follows:

$$y(x) = \sum_{i=1}^{n}(a_i - a_i^*)K(x_i, x) + b \tag{4}$$

In this study,the wavelet kernel function is used as the kernel function of support vector machine,the wavelet kernel function can be shown as follows:

$$k(x, x') = \prod_{i=1}^{M} \mathrm{H}\left(\frac{x_i - b}{a_i}\right)\mathrm{H}\left(\frac{x_i' - b'}{a_i}\right) \tag{5}$$

where x_i denotes the variable vector.

3.Experimental Testing and Analysis

As shown in Fig.1,aviation accidents from 1995 to 2003 in China are employed as our experimental data to test and analysis for aviation accidents evaluation by using the proposed method.The two dimensional input vector of the training samples is employed to construct the training samples.The traditional radial basis function kernel-based support vector machine is used to compare with the wavelet kernel-based support vector machine.As shown in Fig.2,the evaluation curve of the wavelet kernel-based support vector machine is given,the evaluation curve of the wavelet kernel-based support vector machine is near to the curve composed of the actual values. As shown in Fig.3,the evaluation error of the wavelet kernel-based support vector machine belongs to the range from 0.015 to 0.04.

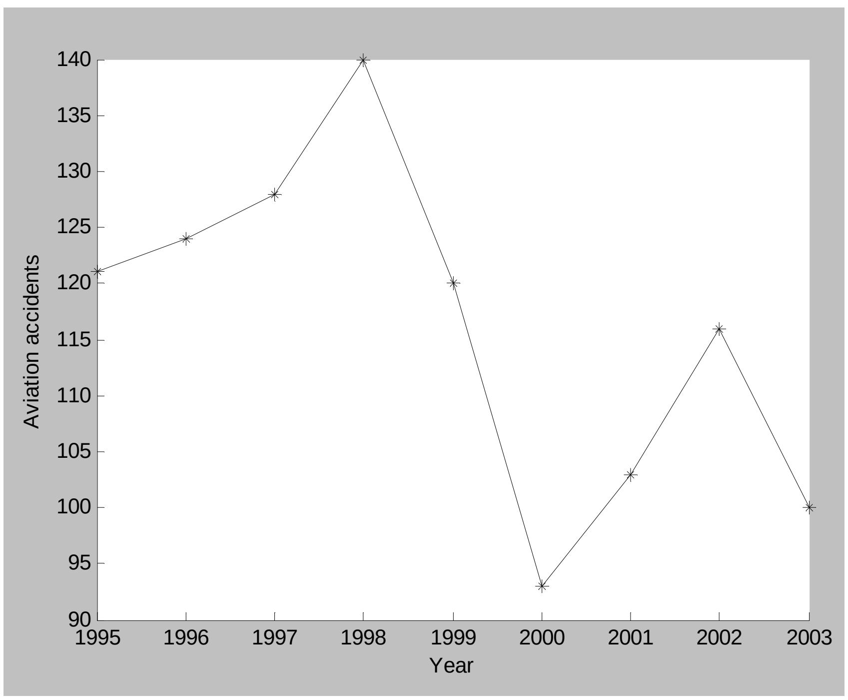


Figure 1. Aviation accidents from 1995 to 2003 in China

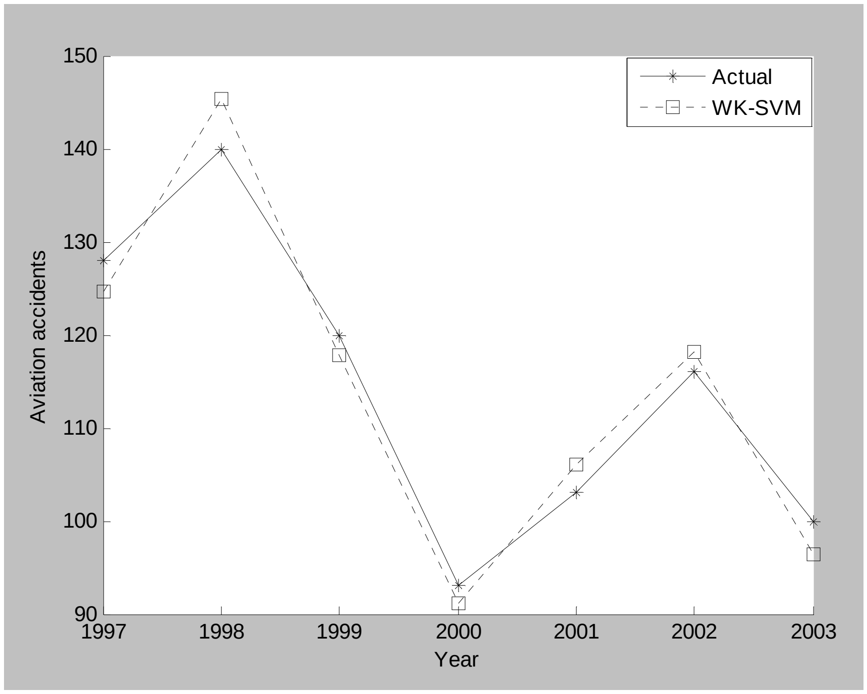


Figure 2. The evaluation curve of the wavelet kernel-based support vector machine

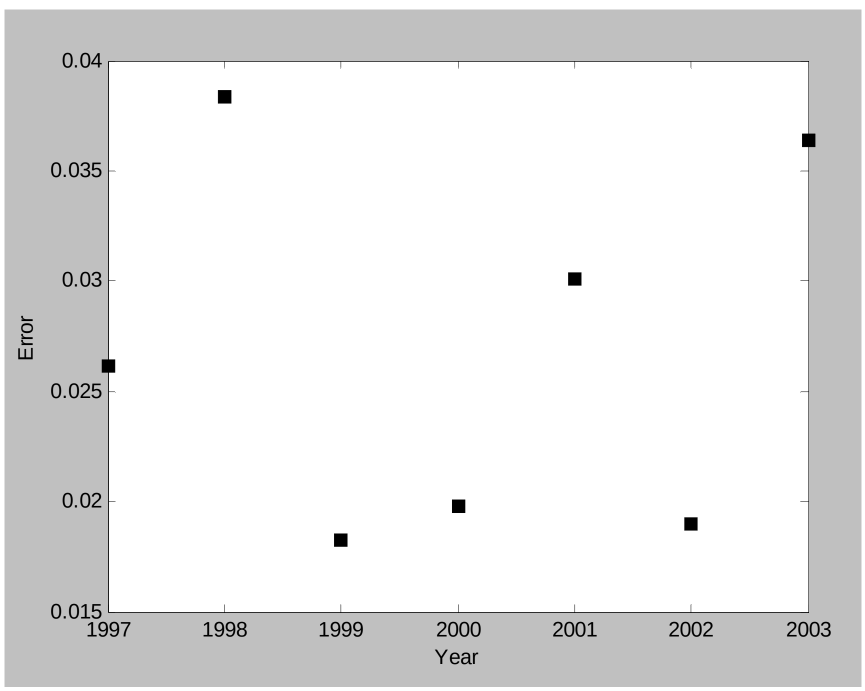


Figure 3. The evaluation error of the wavelet kernel-based support vector machine

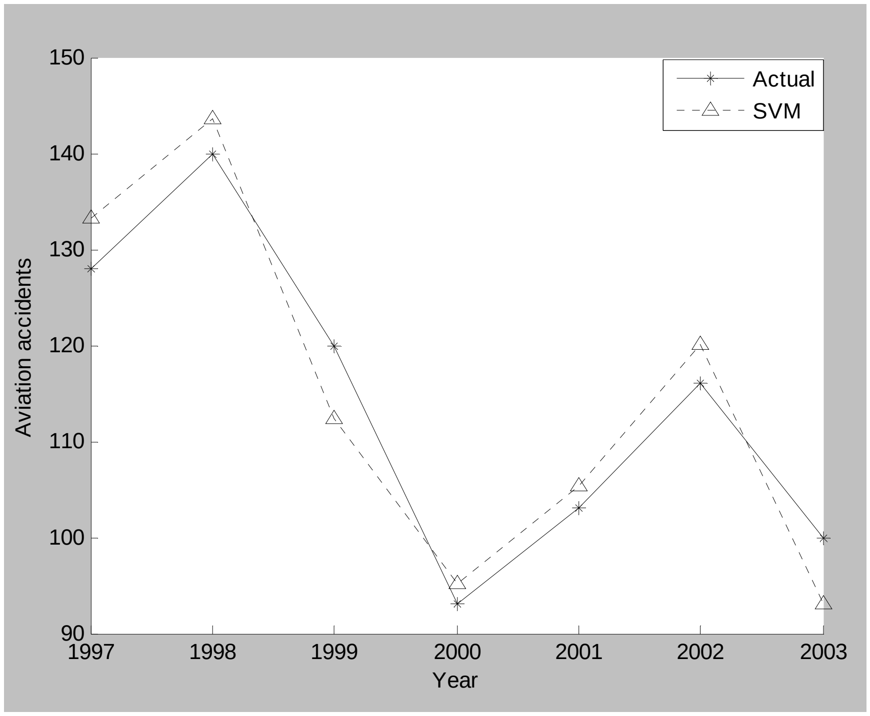


Figure 4. The evaluation curve of the traditional radial basis function kernel-based support vector machine

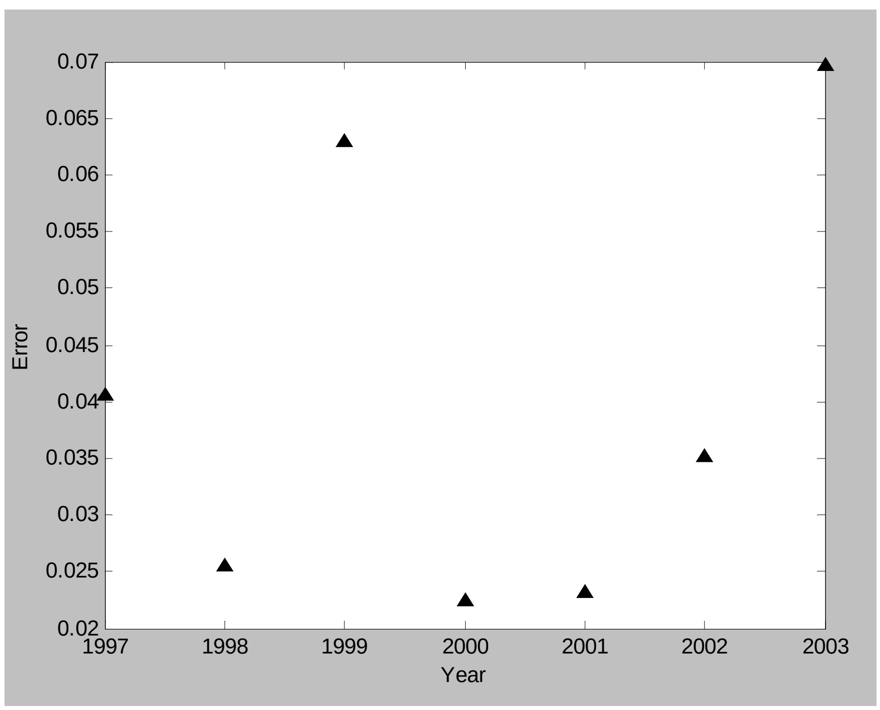


Figure 5. The evaluation error of the traditional radial basis function kernel-based support vector machine

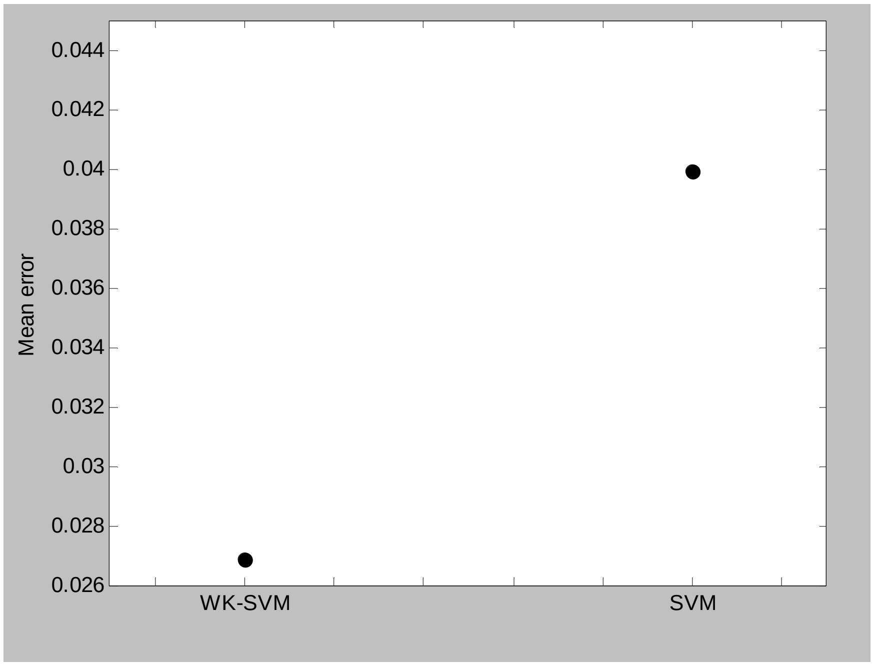


Figure 6. The comparison of the mean error between traditional radial basis function kernel-based support vector machine and wavelet kernel-based support vector machine

As shown in Fig.4,the evaluation curve of the traditional radial basis function kernel-based support vector machine is given,the evaluation curve of the traditional radial basis function kernel-based support vector machine is near to the curve composed of the actual values. As shown in Fig.5,the evaluation error of the traditional radial basis function kernel-based support vector machine belongs to the range from 0.02 to 0.07.

As shown in Fig.6,the comparison of the mean error between traditional radial basis function kernel-based support vector machine and wavelet kernel-based support vector machine is given,which indicates that aviation accidents evaluation accuracy of wavelet kernel-based support vector machine is higher than those of traditional radial basis function kernel-based support vector machine.

4.Conclusion

Wavelet kernel-based support vector machine is applied to aviation safety evaluation. The wavelet kernel function is used as the kernel function of support vector machine,and the two dimensional input vector of the training samples is employed to construct the training samples.The traditional radial basis function kernel-based support vector machine is used to compare with the wavelet kernel-based support vector machine.The testing results indicate that aviation accidents evaluation accuracy of wavelet kernel-based support vector machine is higher than those of traditional radial basis function kernel-based support vector machine.

References

[1] Vahid Khatibi.B, Dayang N. A. Jawawi, Siti Zaiton Mohd Hashim, Elham Khatibi, "Neural Networks for Accurate Estimation of Software Metrics", IJACT, Vol. 3, No. 10, pp. 54 ~ 66, 2011.

[2] Pierre E. ABI-CHAR, Bachar EL-HASSAN, Abdallah MHAMED, "An Enhanced Authenticated Key Agreement Protocol with a Neural Network-based Model for Joining-Phase in Mobile Environments", IJEI, Vol. 2, No. 2, pp. 103 ~ 112, 2011.

[3] K Lamamra, K Belarbi, "Comparison of Neural Networks and Fuzzy Logic Control Designed by Multi-objective Genetic aWlgorithm", IJACT, Vol. 3, No. 4, pp. 137 ~ 143, 2011.

[4] Luo Shihua, Jia Li, "Silicon Content Prediction Using the Hybrid Model by Fuzzy C-means Clustering and Artificial Neural Networks", AISS, Vol. 3, No. 8, pp. 78 ~ 84, 2011.

[5] Jinxue Sui, , Li Yang, Zhilin Zhu, "Flame Burning Condition Recognition Based on Improved ART Neural Network", IJACT, Vol. 3, No. 4, pp. 32 ~ 40, 2011.

[6] Ignacio Yélamos, Gerard Escudero, Moisès Graells, Luis Puigjaner, "Performance assessment of a novel fault diagnosis system based on support vector machines", Computers & Chemical Engineering, Vol. 33, No. 1, pp.244~ 255, 2009.

[7] Haldun Aytuğ, Serpil Sayın, "Using support vector machines to learn the efficient set in multiple objective discrete optimization ",European Journal of Operational Research, Vol.193, No. 2, pp.510~519,2009.

[8] Enrique Romero, René Alquézar, "Comparing error minimized extreme learning machines and support vector sequential feed-forward neural networks",Neural Networks, Vol. 25, pp.122 ~129,2012.

[9] Qian WEI, Jingsha HE, , Xing ZHANG, "A User Admission Scheme for Civil Aviation Emergency Mesh Networks ", JDCTA, Vol. 5, No. 6, pp. 275 ~ 281, 2011.

[10] Tony Bellotti, Roman Matousek, Chris Stewart,"A note comparing support vector machines and ordered choice models' predictions of international banks' ratings",Decision Support Systems, Vol. 51, No.3, pp.682~687,2011.

[11] N. Shojai Kaveh, F. Mohammadi, S.N. Ashrafizadeh, "Prediction of cell voltage and current efficiency in a lab scale chlor-alkali membrane cell based on support vector machines ", Chemical Engineering Journal, Vol. 147, No.2-3, pp.161 ~172, 2009.

COLLABORATION, COORDINATION AND INFORMATION REQUIREMENTS FOR THE SUPPORT OF AN AIRPORT DEPARTURE METERING PROGRAM

Philip J. Smith, Kristen Weaver, Alicia Fernandes, Amy Spencer, Mark Evans and Ken Durham
Department of Integrated Systems Engineering, The Ohio State University, Columbus, Ohio

A simulation system called the Collaborative Airport Traffic System (CATS) has been developed to study human factors issues that arise in the design of a departure metering program to manage the inventory delivered to the spots at an airport over time. CATS is designed to support the role of a Departure Reservoir Coordinator (DRC), whose job is to:

- Manage the inventory at the spots so that the departures queues maintain a "reasonable length" (long enough to avoid having the queue run dry; short enough to avoid unnecessarily high fuel consumption and long tarmac times).
- Manage the inventory at the spots to deal with departure fix constraints, avoiding an inventory at the spots with too many flights to a constrained departure fix. As one Tower TMC put it: "We've got to give the Ground Controller a fighting chance".

This paper describes the design features of CATS and highlights conclusions based on previously reported studies involving cognitive walkthroughs with ARTCC, TRACON and airport Tower traffic managers and with airline dispatchers

The FAA Surface Office is currently defining and refining an operational concept for metering of departures in order to reduce the time spent in departure queues. They are also defining and refining the procedures and software necessary to support such departure metering. The fundamental concepts are:

- Rather than applying a first-come first-served process for developing departure queues, flights should be assigned to virtual queues so that they do not need to enter the departure queue for a runway as early.
- Similar to current practice with Ground Delay Programs and Airspace Flow Programs, slots in this virtual departure queue should be assigned based on a form of ration-by-schedule. These slots are defined by spot times. Two primary parameters need to be set by the DRC in order to manage this rationing process: Predicted departure rate for a runway and target departure queue length.
- Given such a virtual queue, on average, flights should be able to remain at the gate longer without the need to board passengers as early, thus reducing Out-to-Off times.
- When a gate is needed for an arrival, the departure may have to push back earlier than necessary to meet its assigned target spot time, but may still be able to wait in some staging area with its engines off or with only one engine on, thus reducing fuel burn.
- Within this virtual queue, a flight operator should be allowed to swap its flights in order to expedite the departure of an "important" flight.

In our work, we are focused on the following issues:

- The allocation of roles and responsibilities.
- The support of coordination and collaboration across different facilities and organizations.
- The definition of procedures.
- Information and information exchange requirements.
- Functional requirements for supporting technologies.
- Interface design requirements.

Below, we discuss the design of CATS and, where appropriate, refer to findings from previously reported cognitive walkthroughs. These two cognitive walkthroughs included traffic managers, dispatchers and ramp controllers from the following facilities:

- Towers: JFK, LGA, EWR, IAH, DTW, MEM, PHL, MEM, ORD, LAX, MIA, DFW, SFO
- TRACONS: New York, Detroit, SoCal, Dallas
- ARTCCs: ZNY, ZOB, ZME, ZFW, ZMP; ZKC, ZOA
- Ramp Control: JFK – DAL, JBU; EWR – UAL; MEM – FedEx
- AOCs: JBU, AAL, FedEx

- Port Authority: JFK.

Discussion of the Capabilities Embedded in CATS

CATS is designed to support the role of the DRC, and consists of a set of linked displays, including:

- A Selection Tool for highlighting collections of flights that share some property (such as all flights filed to depart a given departure fix). When such a group of flights is highlighted in the Selection Tool, those flights are also highlighted in a Flight List and on a Surface Map.
- The Flight List shows the flights scheduled to depart, along with a number of properties for each flight (such as its filed departure fix and destination).
- The Surface Map shows the location of flights on the airport surface. In terms of the usefulness of the Surface Map, one Tower traffic manager noted: "With this tool, I won't have to get up all the time and look out to see where the flights to a restricted fix are located", (Smith, et al., 2012).
- The Weather Map shows actual and forecast CWIS weather relative to the departure fixes.
- The Control Chart shows actual departure queue length relative to the target departure queue length over time.

It should be noted that CATS has an underlying stochastic model so that the times for pushback, starting the engines, proceeding to taxi, and off (relative to the time when a flight is cleared to depart) are all random variables in order to simulate the variability that routinely occurs during departures.

The displays found in CATS are presented below. Figure 1 shows the basic display layout. The Selection Tool (left window) shows flights organized by departure gate (North, West, South and East) and departure fix (KNAVE, NELYN, etc.).

- Note that next to each departure fix is a number indicating the number of flights that are either currently departing, active (proceeding to the departure runway under ATC control, in the ramp area (proceeding to depart but still under ramp control) or are expected to push back and depart within the next 60 minutes in order to meet their target spot times.
- The controls at the bottom of the Selection Tool allow the user to determine whether to include the flights in all of these different states in the computation of these numbers, as well as to determine which flights will be highlighted in the Flight List and on the Surface Map when a category is highlighted in the Selection Tool.
- The Flight List provides access to a large number of parameters relevant to a given flight. (Only a subset is shown in Figure 1.)
- The yellow markers in the Flight List indicate flights that are within the "Static Time Horizon" or freeze time, meaning that, in order to give the flight operators some level of stability, their spot times should normally be left unchanged if the need arises to modify target spot times. (Input from SMEs regarding the Static Time Horizon has indicated that it is likely to be set at around 15-30 minutes before pushback.) Observation of SMEs playing the roles of ARTCC and TRACON traffic managers and the DRC use these markers to see the level of demand for a weather impacted fix that they can't affect by changing the metering program to change the inventory delivered to the spots.
- The Surface Map includes an embedded table that indicates the number of actual flights that were off in the last 15 minutes relative to the expected departure rate, and the actual number of flights that are active at the current time relative to the target number of active flights (a surrogate for the length of the departure queue).
- The Surface Map further indicates runway by color (blue aircraft are expected to depart 18C; purple aircraft are expected to depart 18L) and aircraft status (triangles represent aircraft that have pushed back; dots represent aircraft that are at their gates).

Figure 2 shows flights to 4 different departure fixes highlighted. This highlighting of 4 fixes in two colors was used by a pair of ARTCC and TRACON traffic managers to evaluate the impact of putting two MIT restrictions in effect to deal with weather constraints (12 MIT off the ground for WICKR and WILEY as one and 12 MIT off the ground for WORTH and WYMON as one). Our studies have shown that the traffic managers representing ARTCCs, TRACONS and Towers, along with the DRC, make extensive use of the Selection Tool in order to efficiently determine the predicted demand for weather constrained departure fixes (Smith, et al., 2012).

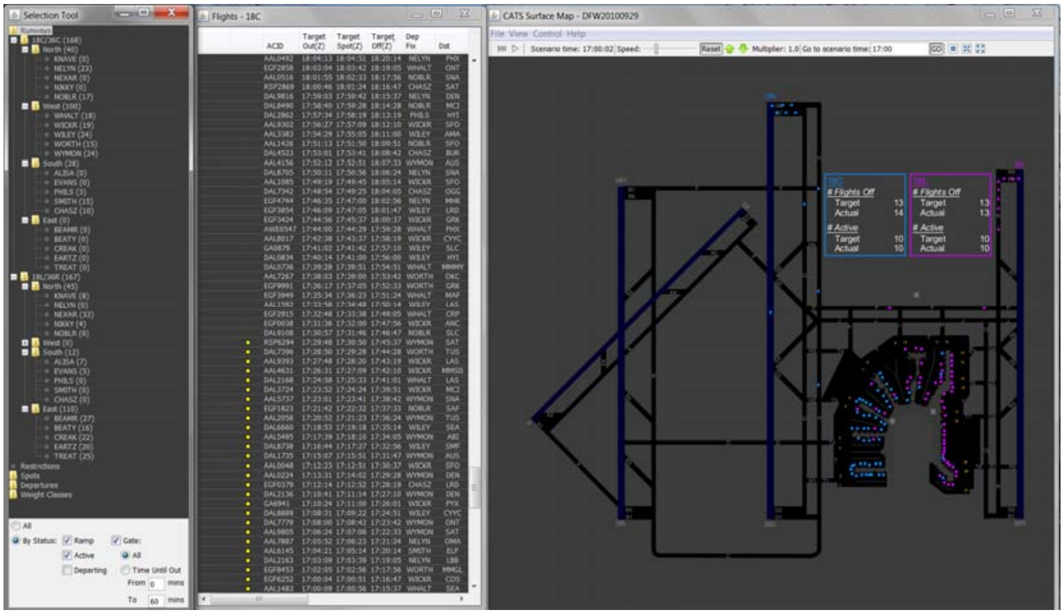


Figure 1. Selection Tool (left window), Flight List (center window) and Surface Map (right window).

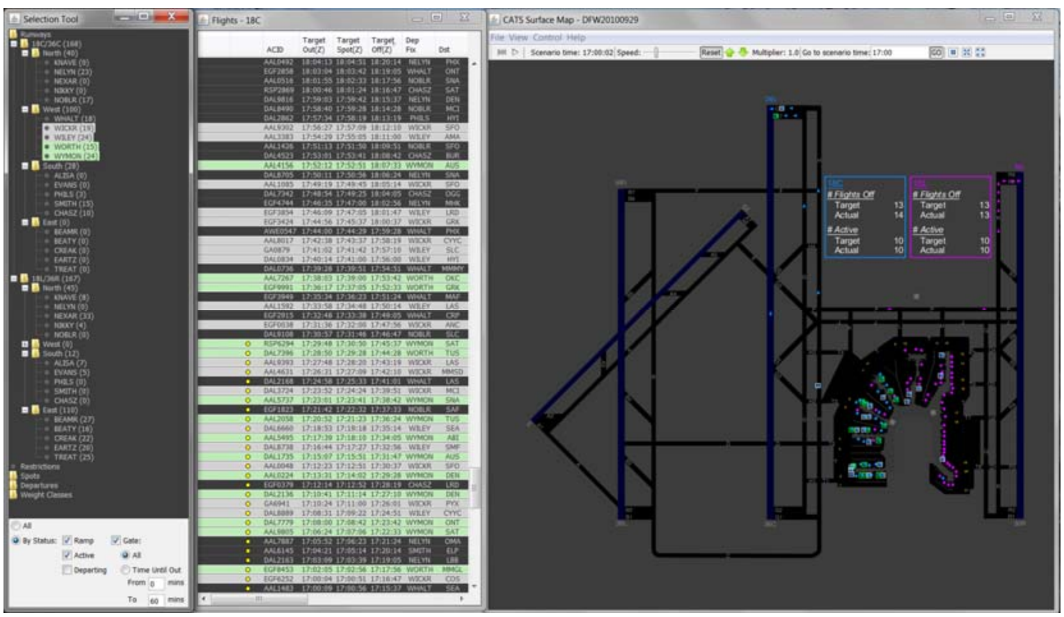


Figure 2. Flights filed to depart via WICKR and WILEY highlighted in gray; flights filed to depart via WORTH and WILEY highlighted in green.

Figure 3 shows the use of the Selection Tool after the two MIT restriction have been put in place (12 MIT off the ground for WICKR and WILEY as one; 12 MIT off the ground for WORTH and WYMON as one). Note that this has been done by highlighting these two active restrictions in the Selection Tool. Note that in this figure we have also introduced the weather display:

- The weather in the pane to the left of the Weather Map window shows the current actual CWIS weather.
- The weather in the pane to the left of the Weather Map window shows the forecast CWIS weather. (This CWIS weather data which is integrated into CATS is for a summer day in ZFW and was provided by MIT Lincoln Labs.)

- The forecast weather can be moved forward in time up to 2 hours in the future, so the user can watch the forecast progression of the weather.
- The actual weather can be moved back in time (a feature requested to make shift change handovers more effective as well as to assess past performance).

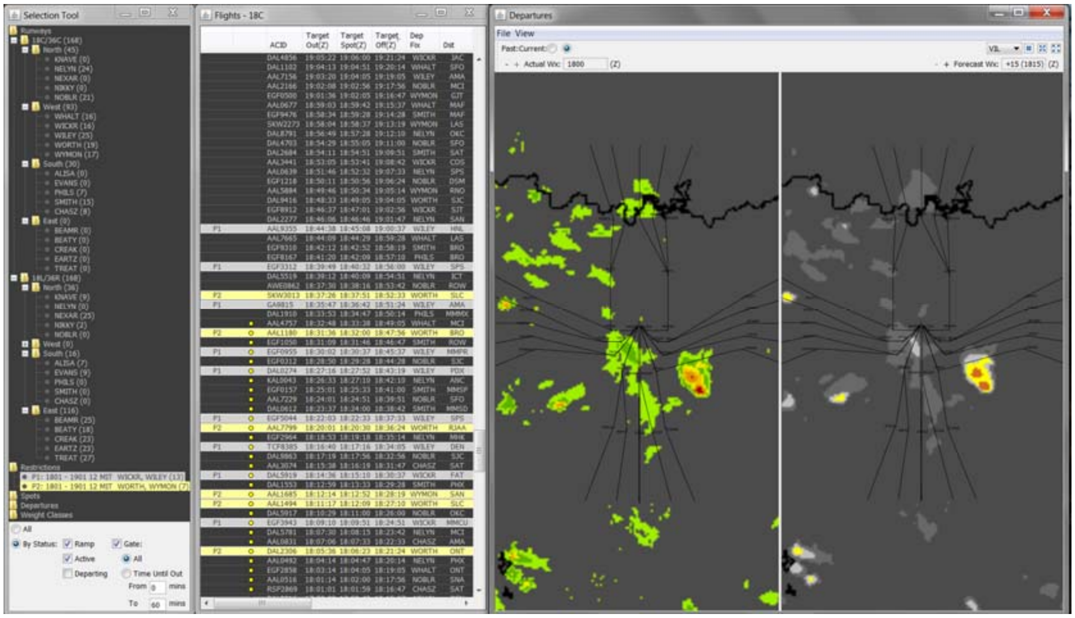


Figure 3. Use of the Selection Tool to highlight the flights affected by two different MIT restrictions.

Figure 4 shows the Control Chart. In CATS, we are using this to monitor past performance. The FAA Surface Office, however, is using a similar chart to predict future departure performance and guide decisions about whether a departure metering program is needed.

Figure 5 illustrates a situation where there is a need for the DRC to manage the inventory at the spots. A 12 MIT restriction off the ground has been put in place for WICKR and WILEY, and the Flight Table shows that there will be a large inventory of flights to these fixes delivered to the spots in a short time period if nothing is done to insert splitters into the spot delivery plan. Figure 6 shows the swapping of flights to insert such splitters, thus "giving the ground controller a fighting chance". CATS automatically made these swaps upon request by the DRC, thus eliminating the need for the DRC to make numerous manual swaps to achieve this reviews plan for delivering inventory to the spots.

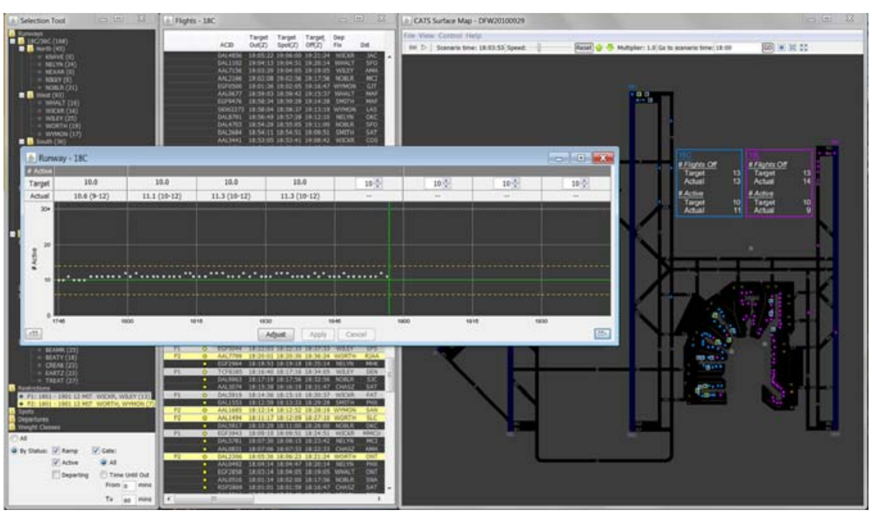


Figure 4. Control chart showing historical performance. The horizontal green line indicates the target number of

active flights; the vertical green line indicates the current time.

Important Human Factors Considerations in Using CATS

Figures 1-6 show an illustrative design of a tool to support the management of departure metering plans by the DRC. Previous studies (Smith, et al., 2012) have highlighted a number of important design considerations, which are summarized here:

- The use of the Selection Tool to highlight flights based on filed departure fix or inclusion in a MIT restriction is of considerable value to ARTCC, TRACON and Tower traffic managers, as well as airline ATC coordinators and dispatchers and the DRC. This tool facilitates individual situation awareness. It also facilitates shared situation awareness and collaboration among the ARTCC, TRACON and Tower traffic managers and the DRC.
- In terms of some specific features of these displays
    - Provide a selection/highlight function that makes it easy to see at a glance a collection of flights based on departure fix, restriction, etc.
        - View by departure fixes or restrictions
        - Link with Flight Table and Surface Map
        - Support at least 2 colors for highlighting
        - Use a background color that provides high contrast with a number of other colors to be used for objects and for highlighting
- Provide a control to determine the flights that are highlighted based on their status (Departing, Active, Ramp, at Gate – Based on Time before TOBT).
- Provide access to a number of flight parameters in the Flight Table (such as weight class and assigned spot), but give the user the ability to order the columns (parameters shown) and sort by different columns.
- ARTCC and TRACON weather restrictions get translated into an airport plan, which should in turn determine the strategy for feeding inventory to the spots in a departure metering program. Thus, inventory needs to be managed to support the airport taxi plan, not just the airspace plan. This makes it important to ensure that the DRC is in the loop regarding the ground controller strategy for feeding the departure queues. (Keeping in mind that there are other considerations as well, this finding supports the concept that the DRC should be an FAA function with the DRC in the Tower.)
- Support a variety of strategies to deal with dynamic convective weather constraints (developing a plan for assigning flights to departure queues by departure fix, inserting splitters to deal with departure fix constraints, change the inventory delivered at the spots to support ground controllers strategies, reroute departures, move flights to holding areas). Different strategies are determined by different individuals. The ARTCC and TRACON traffic managers, for instance, determine what MIT restrictions to put into place, while the ground controller determines how to move aircraft into different departure queues based on departure fix, while the DRC decides how to deliver the appropriate inventory to the ground controller. Thus, coordination between these individuals is critical to effectively manage departures.

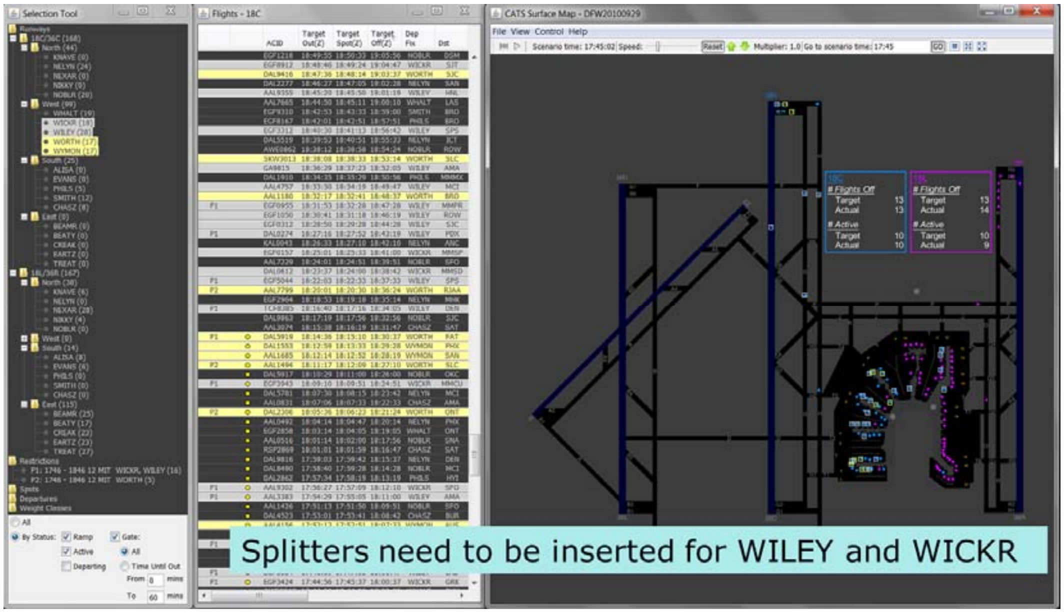


Figure 5. Original plan to deliver WICKR and WILEY flights (highlighted in gray) to the ground controller.

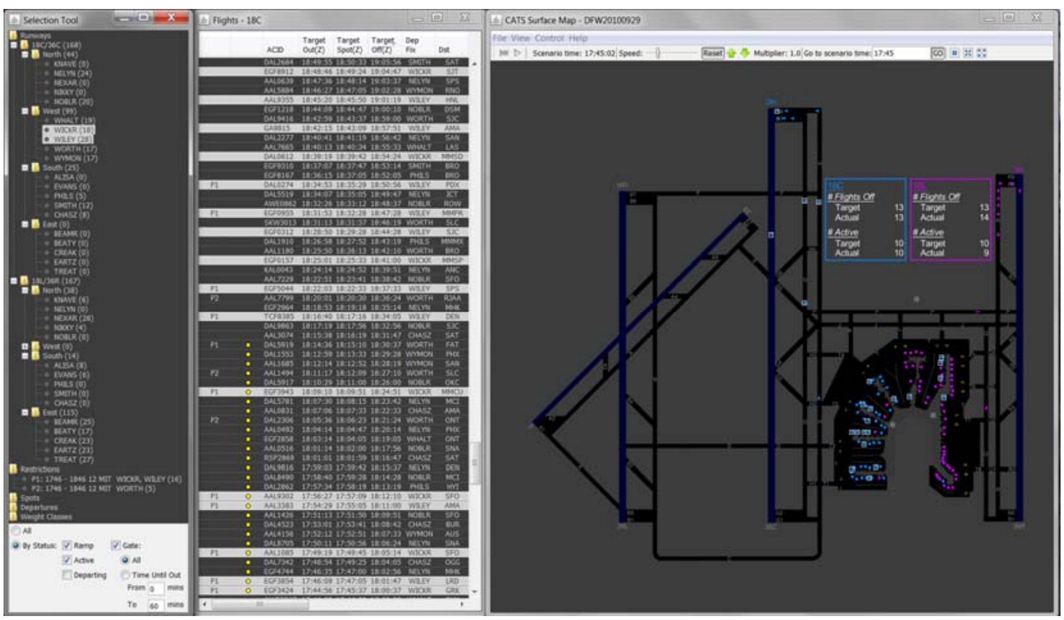


Figure 6. Plan to deliver WICKR and WILEY flights to ground controller after CATS has inserted splitters.

Acknowledgements

The FAA Human Factors Research & Engineering Group coordinated the research requirement and its principal representative acquired, funded, and technically managed execution of the research service. CWIS data was provided by MIT Lincoln Labs.

References

Smith, P.J., Weaver, K., Fernandes, A. Durham, K., Evans, M. Spencer, A. and Johnson, D. (2012). Supporting distributed management on the airport surface. Proceedings of the 2012 Digital Avionics Systems Conference, Williamsburg, VA.

FLIGHT-DECK TECHNOLOGIES TO ENABLE NEXTGEN LOW VISIBILITY SURFACE OPERATIONS

Lawrence (Lance) J. Prinzel III[1], Jarvis (Trey) J. Arthur[1], Lynda J. Kramer[1], Robert M. Norman[2], Randall E. Bailey[1], Denise R Jones[1], Jerry R. Karwac Jr.[3], Kevin J. Shelton[1], Kyle K.E. Ellis[1]

NASA Langley Research Center[1] Hampton, VA; Boeing Commercial Aircraft[2] Seattle, WA; Unisys Corporation[3] Hampton, VA

Many key capabilities are being identified to enable Next Generation Air Transportation System (NextGen), including the concept of Equivalent Visual Operations (EVO) – replicating the capacity and safety of today's visual flight rules (VFR) in all-weather conditions. NASA is striving to develop the technologies and knowledge to enable EVO and to extend EVO towards a "Better-Than-Visual" operational concept. This operational concept envisions an 'equivalent visual' paradigm where an electronic means provides sufficient visual references of the external world and other required flight references on flight deck displays that enable Visual Flight Rules (VFR)-like operational tempos while maintaining and improving safety of VFR while using VFR-like procedures in all-weather conditions. The Langley Research Center (LaRC) has recently completed preliminary research on flight deck technologies for low visibility surface operations. The work assessed the potential of enhanced vision and airport moving map displays to achieve equivalent levels of safety and performance to existing low visibility operational requirements. The work has the potential to better enable NextGen by perhaps providing an "operational credit" for conducting safe low visibility surface operations by use of the flight deck technologies.

NASA is conducting research, development, test, and evaluation of flight deck display technologies that may significantly enhance the flight crew's situation awareness, enable new operating concepts, and reduce the potential for incidents/accidents for terminal area and surface operations. The technologies that form the backbone of the BTV operational concept include: surface and airport moving maps; head-up and head-worn displays; four dimensional trajectory (4DT) guidance algorithms; digital data-link communications; synthetic and enhanced vision technologies; and traffic conflict detection and alerting systems (Bailey, Prinzel, Young, and Kramer, 2011; Prinzel et al., 2011). Preliminary research is described assessing a subset of these technologies in comparison to current-day low visibility surface operations.

The Problem

Research and experience has shown that reduced operational tempos and delays in current-day surface operations due to low visibility conditions contribute significantly, and are growing in their contribution, to airspace delays. During low visibility conditions, pilots and vehicle operators must maintain their situation awareness to ensure the continuation of safe, efficient ground operations. FAA 2010 Annual Runway Safety Report statistics showed that 951 runway incursion events with 12 serious incidents occurred during 52,928,316 surface operations. Although the total number of runway incursions is a very small percentage of total operations, a runway incursion can have catastrophic consequences. The largest category of causal factors in these events was pilot deviations (63%) suggesting that enhancement of situation awareness (ownship position and routing) could provide significant reductions in runway incursions.

Ground-Based Solutions

As a counter-measure in low visibility conditions, the FAA has established regulations, standards, and supporting advisory material in the development of Surface Movement Guidance and Control System (SMGCS) requirements where scheduled Air Carriers are authorized to conduct operations is less than 1,200 feet visibility. SMGCS involves surveillance, routing, guidance, and control for controllers, pilots, vehicle drivers, and other airfield service providers. Key enabling elements of SMGCS are enhanced visual aids – consisting of lights, markings, and signage – designed to provide visual cues for ownship position identification, navigation/route information, and status information for runways, taxiways, hold lines, maneuvering areas, etc.

Flight-Deck Solutions

The low visibility operations (LVO)/SMGCS enhanced visual aids are an established means of creating improved awareness for the crew to ensure the continuation of safe, efficient ground operations.. The present paper describes a "proof-of-concept" test to evaluate the feasibility and efficacy of a flight deck-based approach toward this same objective, specifically using: (a) enhanced vision technology displayed on a HUD and head-down display, and (b) Airport Moving Map (AMM) displays.

These technologies potentially create:

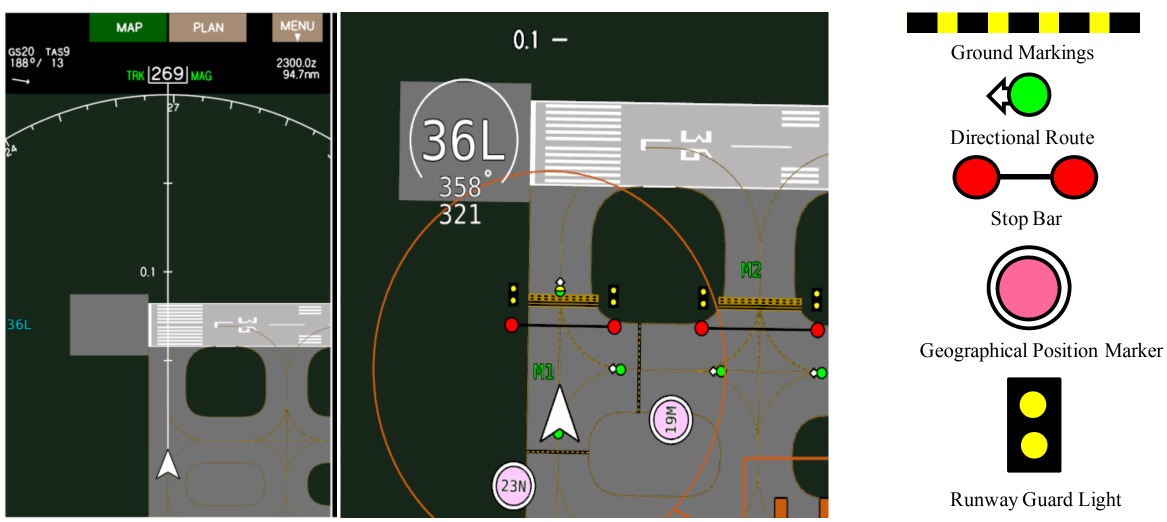
- Improved crew visibility of the airport (topography, surface, and traffic/objects) in the vicinity of the aircraft by an electronic means of enhancing a pilot's natural vision.
- Improved surface position and airport surface status (and also possibly, traffic and object) awareness through airport maps/mapping products (e.g., electronic AMMs).

Enhanced Vision (EV) is an electronic means to provide a display of the external scene by use of an imaging sensor, such as a Forward-Looking Infra-Red (FLIR) or millimeter wave radar. In most atmospheric conditions, especially when natural visibility is reduced due to night, smoke, or haze, the EV provides a visibility improvement which may enable the flight crew (pilot) to more safely operate on the surface. Such a goal is supported by the FAA's mid-term vision for NextGen and an operational improvement (OI) to utilize EV in lieu of SMGSC infrastructure requirements (OI 103208). Over 1000 EV systems are currently flying in the US National Air Space (NAS).

The Commercial Aviation Safety Team (CAST) recommended the use of AMM displays as a highly effective safety enhancement to reduce the risk of runway incursions. Research has supported the conclusion that situation awareness is substantially enhanced by the presence of AMM display that, as a minimum, depicts ownship position. For example, Hooey & Foyle (2007) found that 17% of low visibility and night taxi trials resulted in navigation errors that were mitigated by the use of AMMs. The NTSB has recommended the adoption of AMMs and they are standard equipage on most new commercial transport aircraft.

E-SMGCS Display Concept

The use of AMMs to enhance situation awareness has long been established. However, little research exists that investigated their use under low visibility surface operations. To date, research has been limited to visibility conditions greater than 700 RVR and without the enhanced visual aids required under LVO/SMGCS operations.



Airport Moving Map *E-SMGCS (zoomed in for detail)* *Select E-SMGCS Symbology*

Figure 1. Airport Moving Map (Left), E-SMGCS (Center), Select E-SMGCS Symbology (Right)

In this research, the AMM design was based on NASA "best practices" for AMMs, existing published standards, and an AMM information priority survey. The survey was collected on twenty commercial pilots, experienced with SMGCS and AMMs, based on established methods (Schveneveldt, et al., 2000; Yeh & Chandra, 2003). The results evinced that pilots desired the following SMGCS elements be depicted under low visibility surface operations: Geographic Position Markers (GPMs), clear route, stop bars, and hold lines. These additional informational elements are added to the AMM to create an E-SMGCS display concept.

The E-SMGCS concept is a display mode of the AMM that is invoked by the flight crews when known LVO/SMGCS conditions exists (i.e., under 1200 RVR). The AMM then shows specific LVO/SMGCS information elements based on information priority survey. The E-SMGCS mode retains all normal AMM functionality and also provides the pilots with specific symbology to enable the pilots to cognitively map task priority elements between the AMM, paper charts, and out-the-window visual cues.

The E-SMGCS concept is envisioned to enable the various information elements to be dynamically controlled (e.g., remove stop bar depiction when cleared onto active runway). The E-SMGCS mode may also include display of ownship route and other traffic and their *intended* route (see Prinzel et al., 2010) with 4DT depictions and conflict detection and resolution alerting and indications (e.g., see RTCA DO-323).

Research

A small-scale study was conducted to evaluate the preliminary concept of operation and display for the E-SMGCS mode that is utilized during low visibility surface conditions. Flight crews conducted approach and departure operations using a 6 degree-of-freedom full-motion large glass cockpit simulation (see Figure 2) during 300 ft runway visual range (RVR) low visibility operations at Memphis International Airport (FAA Identifier: KMEM).

The objectives of the research were:

1. Assess the use of EV on HUD (enhanced flight vision system or EFVS) during LVO/SMGCS and its potential flight deck impact when integrated with an AMM with ownship representation.
2. Assess the use of EVFS during LVO/SMGCS and its potential flight deck impact when integrated with an AMM with SMGCS-specific symbology (i.e., the E-SMGCS concept)
3. In evaluating the use of EVFS, consider the possibility of operational credit for use of EV in lieu of airport SMGCS enhanced visual aid equipage.

Four commercial flight crews (Captain, F/O), paired for same airline, served as the subjects. Pilots were HUD-qualified and had EV experience and commercial airline operational experience with SMGCS. The Captain was seated on the left-side and was "pilot-in-command" responsible for approach, landings, and taxi of aircraft. The First-Officer was seated on the right-side and functioned as "monitoring" pilot.

A two (LVO/SMGCS Level 1, LVO/SMGCS Level 2 out-the-window enhanced visual aids) by two (EV, none) by three (none, AMM with ownship, AMM with ownship and E-SMGCS) partially factorial within-subjects factor design was conducted (Figure 2):

- The simulation was carefully constructed to faithfully replicate the SMGCS enhanced visual aids (markings and lights) during Level 1 (1200-500 ft RVR) or Level 2 LVO/SMGCS (>500 ft RVR) at KMEM. For the present research, Level 1 contained taxiway centerline lighting, runway guard lights, edge lighting, and all other requirements in addition to GPMs that functioned as "surface painted location signs." Level 2 retained all of Level 1 SMGCS and additionally, provided controllable stop bars, and clearance bar lights (co-located with GPMs).
- The EV display concept used forward-looking infrared (FLIR) depicted on a wide field-of-view (42^{o} H x 30^{o} V) Head-Up Display (HUD) with a head-down repeater display of the FLIR (no symbology).

Scenarios were constructed using subject matter experts to create representative types of operations and flows typically experienced during low visibility surface operations. 300 ft RVR day time conditions were simulated. The EV (FLIR) provided 500 ft visibility of topology and 700 ft visibility of lights. Scenarios were balanced in presentation and included taxi-out for departure and arrivals with taxi-in to the ramp. All approaches were flown using an auto-land capability. An off-nominal trial was presented which failed the FLIR input to the HUD and head-down repeater display.

Quantitative and qualitative measures were collected during trials including Situation Awareness Rating Technique (SART) and NASA Task Load Index (NASA-TLX). SmartEye™ eye tracking measures were also recorded.

The full-motion simulator (Figure 2) modeled a large commercial transport aircraft with typical weight and balance (approximately 180,000 pounds gross weight, 25% cg using 30,000 pounds fuel) and is configured to mimic the instrument panel of current state-of-the-art aircraft, with four 10.5" Vertical (V) by 13.25" Horizontal (H), 1280x1024 pixel resolution color displays tiled across the instrument panel. A collimated out-the-window (OTW) scene provided approximately 200 degrees horizontal by 40 degrees vertical FOV at 26 pixels per degree.

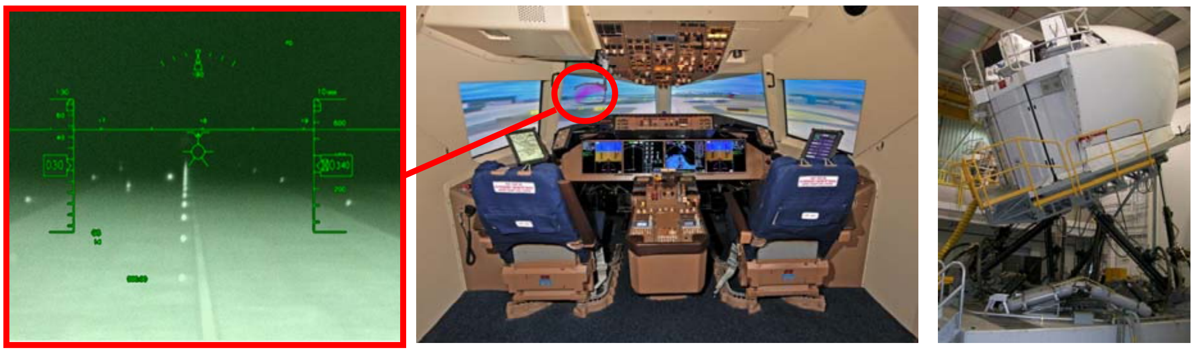


Figure 2. Full Motion-Based Commercial Aircraft Simulator (Enhanced Vision HUD shown)

Briefings and simulator training was provided to ensure familiarity with KMEM and simulation set-up. Pilots were instructed to conduct operations that reflected the standard operating procedures and communications during LVO/SMGCS operations. All radio communications were pre-recorded but were backed-up with a live controller confederate. Flight crews were briefed to conduct the operation as though passengers were onboard and to emphasize aircraft safety and comfort common for line operations. A total of nine experimental trials (8 nominal, 1 off-nominal) were conducted.

Results and Discussion

Data analyses are on-going; not all results were not available at publication (e.g., eye tracking data are not reported).

Quantitative taxi performance data analyses have been conducted and no significant effects were found, $p > 0.05$. Average taxi speed was 10.69 knots and was consistent across display and SMGCS conditions and reflects typical taxi speed under visibility conditions.

Situation Awareness. A significant main effect was found for display condition for situation awareness, SA (i.e., using SART), $F(3,18) = 51.958$, $p < 0.0001$; and SMGCS condition, $F(1, 6) = 16.754$, $p < 0.01$. No significant differences were found between pilot role of Captain (SART = 3.594) and First Officer (SART = 3.250), $p > 0.05$.

Flight crews reported higher SA under the Level 2 SMGCS (3.938) compared to Level 1 SMGCS (2.906). Pilots also reported significantly higher SA using the EV+E-SMGCS AMM (7.313) compared to the other three display conditions: No EV+No AMM (0.250), EV+No AMM (2.125), EV+AMM-Ownship (4.000). The post-hoc tests also evinced that having EV was significantly better than not having EV but that the addition of the AMM display significantly enhanced SA further compared to EV+No AMM condition. No interaction was found between SMGCS x Display Type, $p > 0.05$. Post-test paired comparison SA results reflected the SART results and pilots rated the EV+E-SMGCS display to be significantly better for SA compared to the other three display conditions (in rank order): EV+AMM-Ownship, EV+No AMM, or No EV+No AMM display concepts tested. The EV+AMM-Ownship condition was also significantly rated higher in SA than EV+No AMM or No EV+No AMM conditions.

Mental Workload. The TLX ratings analysis revealed a significant main effect for display condition, $F(3, 18) = 166.8$, $p < 0.0001$ and a trend for significance in the SMGCS x Display interaction, $F(1,6) = 2.514$, $p = 0.91$ which, given the small N, is reported here for consideration. The results demonstrated that pilots reported the EV+E-

SMGCS (20.625) to be significantly lower in mental workload than EV+AMM-Ownship (35.00), EV+No AMM (73.750) or No EV+No AMM (75.625). Although not significant, the interaction for SMGCS x Display purports that pilots reported lower workload when using Level 2 LVO/SMGCS visual aids only for the display conditions that had an AMM; pilots reported slightly higher mental workload when using Level 2 LVO/SMGCS under the non-AMM display conditions.

Post-Run Questionnaire. Ten post-run questions were administered. All questionnaire items were found to be significant, $p < 0.05$. Table 1 contains the means for the questionnaire by display condition.

Table 1: Post-Run Questionnaire Means

	Level 1 LVO/SMGCS				**Level 2 LVO/SMGCS**			
	NO EV + NO AMM	*EV + NO AMM*	*EV + AMM*	*EV + E-SMGCS*	*NO EV + NO AMM*	*EV + NO AMM*	*EV + AMM*	*EV + E-SMGCS*
Q1	4.88	4.63	1.75	1.00	4.25	4.5	1.38	1
Q2	5.00	3.88	2.00	1.00	4.50	4.5	2	1
Q3	3.38	3	1.75	1	3.38	3	1.75	1
Q4	4	3.50	2.75	2.5	4	3	3	2.75
Q5	4	3.5	2.25	2.25	3.75	3.5	2.5	2.5
Q6	N/A	1.75	1.25	1.25	N/A	2.25	1.25	1.25
Q7	3.63	3.5	2	1.75	3.38	3.5	1.75	1.38
Q8	4.38	3.5	2	1	3.38	4.25	1.75	1.25
Q9	N/A	N/A	1.63	1	N/A	N/A	1.75	1
Q10	4.75	3.75	1.5	1.25	4	3	1.25	1

Note: 1 = Strongly Agree; 4 = Neither Agree or Disagree; 7 = Strongly Disagree

Q1. I was able to maintain taxi accuracy during SMGCS operation.
Q2. I was aware of ownship position on the airport surface.
Q3. I was aware of the cleared SMGCS taxi route.
Q4. I was aware of traffic and other vehicles during SMGCS operation.
Q5. I was aware of SMGCS signage, markings, and visual aid.
Q6. The FLIR presentation was effective for SMGCS taxi operation (ease of access, size, etc.).
Q7. The display concepts and SMGCS charts contributed to communication effectiveness (ATC and Flight Crew)
Q8. The display concepts and SMGCS charts promoted crew resource management, coordination, and cohesion.
Q9. The airport moving map (if applicable) display was effective for situation awareness during SMGCS operation.
Q10. The display concepts and SMGCS charts contributed to perceived safety during SMGCS taxi operation.

Off-Nominal Event. During the final trial (unbeknownst to flight crews), the EV (FLIR) failed during taxi at a critical geographical position in which the flight crews needed to first detect the failure and then decide upon the proper course of action given aircraft location and the capabilities available inside and outside the flight deck (all off-nominals were conducted during Level 1 LVO/SMGCS equipage). The failure was presented to pilots as within-subjects variable and the display conditions were EV+No AMM, EV+AMM-Ownship, or EV+E-SMGCS AMM. Pilots reported that they felt significantly safer and were willing to continue the operation (e.g., cross an inactive runway) without the EV if an AMM was available. When the AMM was not present, the flight crew terminated operation and requested to return to gate with assistance. To verify that result, a second flight crew (Crew #4) also was presented with off-nominal scenario with baseline display and, as with the other flight crew, also terminated the operation and requested follow-me vehicle assistance.

When the AMM was available, flight crews significantly reported higher SA with both the AMM-Ownship (5.65) and AMM with E-SMGCS (6.00) compared to the baseline, no AMM condition (1.50). The baseline condition received significantly higher TLX scores (82.00) compared to AMM (35.00) and E-SMGCS (42.00). The location of ownship and labels on moving map were judged to be most significant contribution toward SA followed by depiction of stop bars and runway guard lights and geographical position markers on the E-SMGCS. Pilot consistently rated “strongly agree” that a EV (FLIR on HUD) with an AMM may allow for “operational credit” to reduce ground-based requirements for LVO/SMGCS enhanced visual aids and that the E-SMGCS was judged to be significantly better for SMGCS operations than a basic AMM. Pilots also emphasized the desire for ownship routing and display of other traffic with CD&R alerting as a bonus.

Conclusion

The results demonstrated that an enhanced flight vision system may potentially enhance situation awareness and ameliorate problems witnessed when visibility drops requiring the use of LVO/SMGCS enhanced visual aids. However, the use of EV alone was not found to substantially enhance surface operations compared to baseline (i.e., no FLIR) without the addition of an AMM. Pilots consistently rated the AMM to be of significant value for these operations and, together, the EV and AMM was rated to be of tremendous benefit in maintaining SA and workload during 300 RVR approach and departures with simulated taxi-in and -out. The results also fully support the potential direction that EV with an AMM may provide an "operational credit" for SMGCS wherein an operator, with these requisite flight deck technologies, may be able to conduct lower than 500 RVR operations at airports that may only have a Level 1 LVO/SMGCS airport visual aids in place. Another option may be to enable under 1200 RVR surface operations at airports that do not have any LVO/SMGCS airport visual aids in place.

The FAA has stated that, "taxiing on the airport surface is the most hazardous phase of flight" (Gerold, 2001). Almost a decade later, that statement still rings true, but LVO/SMGCS enhanced visual aids and other controls are significantly improving this situation. Emerging flight deck technologies offer a potential means to create an equivalent level of safety and performance. These flight deck technologies, such as the E-SMGCS -AMM display and EV, could assist in fully realizing the potential of NextGen by offering a more affordable path toward safe and efficient LVO/SMGCS operations through an "equivalent visual" paradigm.

References

Bailey, R.E., Prinzel, L.J., Kramer, L.J., & Young, S.D. (2011). *Concept of Operations for Integrated Intelligent Flight Deck Displays and Decision Support Technologies.* NASA Langley Research Center, NASA/TM-2011-217081, Hampton, VA. Apr 2011.

Gerold, A. *Runway Incursions: The Threat on the Ground. Avionics Today*. http://www.aviationtoday.com/av/commercial/Runway-Incursions-The-Threat-on-the-Ground_12628.html Accessed January 10, 2013.

Hooey, B.L., Foyle, D.C. (2007). Aviation Safety Studies: Taxi Navigation Errors and Synthetic Vision Systems Operations, In D. C. Foyle and B. L. Hooey (Eds.) *Human Performance Modeling in Aviation*. Boca Raton: CRC/Taylor & Francis

Federal Aviation Administration (2011). *Annual Runway Safety Report 2010*. Washington, D.C.: FAA.

Prinzel, L.J., Bailey, R.E., Shelton, K.J., Jones, D.R., Kramer, L.J., Jarvis, J.A., Williams, S.P., Barmore, B.E., Ellis, K.E., & Rehfeld, S.A. (2011). Better-than-visual technologies for next generation air transportation system terminal maneuvering area operations. *International Symposium on Aviation Psychology 2011 Bi-Annual Meeting.*

Schveneveldt, R., Beringer,D., Lamonica, J., Tucker, R., and Nance, C. (2000) . *Priorities, organization, and sources of information accessed by pilots in various phases of flight*. Washington, D.C.: Federal Aviation Administration.

Yeh, M. & Chandra, D. (2003). Air transport pilots' information priorities for surface moving maps. *Proceedings of the Human Factors and Ergonomics 47th Annual Meeting, 47*.

Acknowledgments

The research was supported by the NASA's Aviation Safety Program. The authors would like to acknowledge the contributions of many NASA Langley technicians, programmers, and engineers who support/have supported these research projects. Notable among them are Regina Tober, Lon Kelly, Wei Anderson, Thomas Feigh, Chris Harrison, Victoria Chung, Sean Kenny, Dennis Frasca, Tom Wolters, Joe Whiting, Wayne Burge, Catherine Buttrill, Kemper Kibler, Darrell Sacra, Philip Smith, Sonia Herndon, Dale Ashcom, Ben Lewis, Brian Hutchinson, Lindsey Lowe, Donald Buhl, Brent Weathered, Steve Velotas, and Lisa Rippy.

INDIVIDUAL PILOT FACTORS PREDICT SIMULATED RUNWAY INCURSION OUTCOMES

Kathleen Van Benthem (kathy_vanbenthem@carleton.ca)
Institute of Cognitive Science, Carleton University, 1125 Colonel By Drive
Ottawa, Ontario, K1S 0A3 Canada

Chris M. Herdman (chris_herdman@carleton.ca)
Visualization and Simulation Centre, Carleton University, 1125 Colonel By Drive
Ottawa, Ontario, K1S 0A3 Canada

Runway incursions are a critical issue facing the aviation industry, with general aviation accounting for 77 percent of runway incursions involving pilot deviations. The present study reports on the value of individual pilot factors in predicting the outcome of a simulated surprise runway incursion. Significant predictors of runway incursion management were pilot rating, self-rated awareness of the impact of other aircraft on flight, and perceptions of the mental demands of the flight tasks. In light of the aviation industry's reliance upon pilots' self-monitoring of competence, strategies for reducing runway incursions can capitalize on evidence that higher self-rated ability to maintain a comprehensive mental model of relevant airspace, and better judgments of the mental demands of high workload conditions predicted superior incursion management outcomes.

Runway incursions and excursions are a critical safety issue facing the aviation industry, with costs estimated at a billion US dollars annually (Honeywell Aerospace, 2007). While reducing the occurrence of runway incursions is an obvious first step in dealing with this issue, knowing which pilots might be at risk when encountering an incursion upon final approach is another strategy for dealing with this aviation safety problem. One might think that the ability to deal with unexpected events is regularly assessed in general aviation; however, in accordance with a pilot's rating there is variation in the time between regulated evaluations of aviator competence. The two-year time lapse for private pilots between formal evaluations of competence means that self-monitoring of ability is a critical element in pilot safety. The present study provides a unique analysis of the individual pilot factors of age and experience, and pilot self-rating of situation awareness and mental workload as predictors of simulated unexpected runway incursion outcomes.

Individual Pilot Factors and General Aviation Flight Performance

To the authors' knowledge, no general aviation studies to date have investigated the association of pilot age, experience, and subjective self-ratings of situation awareness and workload with the outcome of a true "surprise" event. In order to explore the likely effects of individual pilot factors on an unexpected incursion event upon final approach, evidence pertaining to other general aviation tasks and individual pilot factors is reviewed.

Age

The issue of age and pilot performance has been studied in longitudinal and cross-sectional analyses from a large cohort of general aviation pilots with consistent findings that younger pilots tend to significantly outperform older pilots on key aviation tasks. (Kennedy, Taylor, Reade & Yesavage, 2010; Taylor, Kennedy, Noda, & Yesavage, 2007; Yesavage, Taylor, Mumenthaler, Noda, & O'Hara, 1999; Yesavage, Jo, Adamson, Kennedy, Noda, Hernandez, Zeitzer, Friedman, Fairchild, Scanlon, Murphy, & Taylor, 2011). Yesavage et al. (1999) reported on a sample of 100 pilots and found that overall simulated flight performance was reduced for the older pilots, with age predicting 22% of the overall variance in performance. In 2000 Taylor et al. reported that while four cognitive factors accounted for 45% of the variance in simulator scores (working memory/processing speed, visual memory, motor coordination, and tracking) age contributed significant variance in addition to the cognitive factors. Taylor et al. (2005; 2007) reported lower performance for older pilots when following air traffic control messages, traffic avoidance, cockpit instrument scanning, and approach and landing ability. In a small sample, low-fidelity simulator study of general aviation pilots, Coffey, Herdman, Brown and Wade (2007) found that older pilots missed significantly more rare critical events (near traffic and cockpit malfunctions) than younger pilots. Coffey et al. suggested that age-effects pertaining to "change detection" might be responsible for the older pilots missing more events both inside and outside the cockpit. In light of age-effects noted for other aviation tasks, some requiring quick decision-making and traffic avoidance, it is posited that pilot age might be a predictor of performance handling an unexpected runway incursion upon final approach.

Flight experience

Taylor et al. (2005) found that older pilots performed worse on a communication task and that higher levels of expertise did not offer protection to the older pilots regarding accuracy of communication performance. Taylor et al. (2007) later found that higher-rated pilots showed fewer declines over time, and expertise offered some protection against declines in communication task performance. Yesavage et al. (2011) also reported that higher rating was associated with better performance for all pilot age groups; however, higher rating did not demonstrate mediation of the age-effects found in this sample. Kennedy, Taylor, Reade & Yesavage (2010) demonstrated that expertise provided an advantage for older pilots on the flight control task of banking performance in a holding pattern. Visual flight rules flight into instrument meteorological conditions is a common pilot violation associated with crashes in general aviation and, as such, has been studied in aviation simulation studies. Causse, Dehais, Arexis, and Pastor (2011) also investigated how pilot experience influenced weather-related go-arounds upon approach and found that pilot experience was a significant predictor of weather-related decision-making. It is proposed, that higher levels of experience, as indexed by rating, will predict better runway incursion management.

Subjective Measures of Workload and Situation Awareness

Subjective Measures of Workload

A pilot's ability to subjectively and accurately rate cognitive workload should reflect an appropriate appreciation of the load caused by current task demands. It is suggested that a pilot who is sensitive to the effects of the task demands should also have a good appraisal of the tasks and the environment in which they take place. A commonly used subjective measure of workload is the NASA TLX Index (Hart and Staveland, 1988). In a review of mental workload in the aviation domain Cain (2007) reported that subjective rating scales have shown some reliability and validity, but that the correlation with objective measures of workload has not been consistent. Vidulich (1988) suggested that subjective measures of workload may be poor predictors of actual performance and describes a series of laboratory experiments where task workload manipulations were not appropriately reflected in subject reports of workload. The present study will explore the utility of subjective workload measures in predicting performance in naturalistic setting where alignment between manipulated and subjective workload might be augmented.

Subjective Measures of Situation Awareness

Situation awareness is best described as the pilot perceiving and integrating relevant stimuli within a meaningful "volume of time and space" and using selected stimuli to build a mental model of the environment and project that model into the future (Endsley, 1988, p. 97). Vidulich, Crabtree, and McCoy (1993) found that a one-item scale representing situation awareness did not adequately represent actual situation awareness performance. SART (Taylor, 1990) is a commonly used three-dimensional self-measure of situation awareness. However SART items, which include rating of task demands and attention supply, have been criticized for reflecting workload constructs rather than situation awareness (Jones, 2000). It is proposed that a scale with multiple items indexing situation awareness pertaining to other aircraft locations and relevancy might be associated with performance handling unexpected events that occur during flight.

Method

Participants

The present study was comprised of a total of 108 pilots. Fifteen pilots had incomplete data for the subjective rating scales and were removed from the current analysis. The remaining 93 pilots were aged 19 to 81 years (*M*=46.5) and held a current medical certification. Pilot expertise included student, private (no additional ratings), private (additional ratings), and advanced (airline transport, commercial, and military) ratings. Table 1 reports the mean age, total flight hours, and recent pilot-in-command hours licensed for each pilot-rating group.

Table 1.
Pilot Rating-Group Characteristics

Rating Group (*N*)	Mean Age (years)	Mean Total Flight Hours*	Mean Pilot-in-Command Hours (past 12 months)*
1. Student (12)	40.1	49.2	6.1
2. Private (no additional rating) (35)	48.8	242.0	17.0
3. Private (additional rating e.g. Instrument Flight Rules) (27)	48.8	681.3	29.7
4. Advanced (ATP, Commercial, Military) (18)	43.3	2038.0	82.2

Note. * denotes significant ANOVA between group differences $p<.05$.

Flight simulator environment

Pilots flew a Cessna 172 non-motion simulator with instruments and controls integrated with Microsoft® Flight Simulator X. Three large screens positioned in an arc in front of the cockpit provided approximately 120 degrees of horizontal and 45 degrees of vertical field of view. Feedback from experienced pilots and flight instructors indicated that the simulator was representative of a Cessna 172 type aircraft.

Procedure

Pilots completed a consent form and a demographic and experience questionnaire before receiving detailed instructions regarding the simulator tasks. After an orientation and practice phase pilots were required to fly three left-hand patterns in a low cognitive workload condition and then three left-hand patterns in a high cognitive workload condition. Only the high workload condition is described as it is the condition of interest in the present study. In the high workload condition the airfield was uncontrolled, the terrain was mountainous, and the pilot interacted with up to four other computer generated aircraft. Pilots were required to provide details of their call sign, aircraft type, and location at routine points during the circuit via radio communication, and note the similar information provided by the "pilots" from relevant computer generated aircraft. The unexpected runway incursion event always occurred at the end of the final pattern in the high workload condition. Several other measures of pilot performance were collected but are not reported on in this analysis.

Measures

Simulated runway incursion outcome measure. The runway incursion outcome measure comprised the last task in the flight simulation protocol. In the final high workload condition the pilot was requested to fly two *touch-and-go* patterns. While the pilot was on final approach a "rogue" simulated aircraft was introduced at the distal end of the runway and traveling in the wrong direction (i.e. towards the oncoming pilot participant). To successfully manage this runway incursion pilots must have accomplished three main tasks in quick succession. The initial requirement was to correctly perceive and interpret the environmental stimuli; the second task was to use the interpreted information in order to make a decision regarding the required procedure to avoid the quickly approaching offending aircraft; finally, the correct decision regarding evasive action in the air must be selected and acted upon in timely and efficient manner. The runway incursion outcome score was based on the quality, timing and result of pilot response to the offending aircraft and could range from 0 (did not notice the incursion) to 10 (noticed the incursion with adequate time to make a radio call and perform a safe maneuver to avoid the incursion with no dangerous loss of separation between aircraft): scores below 5 indicate high risk for a poor outcome.

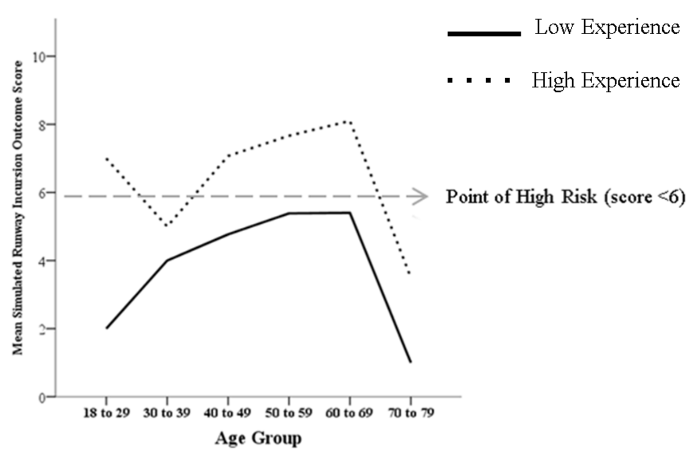
Pilot subjective ratings. A modified (for clarity of instructions and tasks) version of the NASA Task Load Index (TLX) (Hart and Staveland, 1988) was used to gauge pilots' perception of workload in both workload conditions. There were seven categories of workload on a continuous 100mm (end points low and high) rating scale: physical demand, mental demand, temporal demand, performance, effort, and frustration. Subjective situation awareness was measured using a 7-point scale designed specifically for aviation tasks at the ACE Laboratory at Carleton University and was based on the scale first reported by Hou, Kobierski and Brown (2007). The scale consisted of 11 items representing various aspects of situation awareness encountered during flight and included reference to awareness of the environment, tasks, time, priorities, relative position of ownship and other aircraft, future events, and overall level of situation awareness.

Results

Age, Experience, and Runway Incursion Outcome

Figure 1.

Mean Simulated Runway Incursion Outcome by Age and Experience Groups.



Note. Student and private pilot with no additional ratings are Low Experience pilots
Private pilots with additional ratings and ATP, commercial, and military are High Experience pilots

It is clear from Figure 1 that both younger and older pilots performed worse than the middle-aged pilots. Figure 1 suggests that age effects on runway incursion management might be obscured by the tendency of younger pilots to also score poorly on the incursion task. Not surprisingly, there was a marginally significant cubic relationship between age and incursion management, $F(3, 89)=2.43$, $p=.07$. Figure 1 also demonstrates the effect of experience on incursion management scores. Both low and high experience pilots improved their scores after age 30 and then saw a reduction in performance after age 70.

A one-way ANOVA showed a significant overall effect of pilot rating, $F(3, 89)= 9.05$, $p=.001$, $\eta_p^2 = .23$. Figure 1 suggests that highly experienced younger pilots might gain additional benefits for experience not enjoyed by the oldest pilots. However, the age x experience interaction was non-significant, $p>.05$.

Subjective Self-rating of Workload, Situation Awareness, and Runway Incursion Outcome

Pilots who recognized the higher level of mental demand in the high workload condition also tended to better manage the runway incursion. The mental demand item on the workload rating scale was the only item to correlate significantly with the runway incursion outcome.

Table 2.

Significant Predictors of Simulated Runway Incursion Outcome.

Predictor	Linear Regression Coefficients Standardized Beta Coefficient	*t*- value
Pilot Rating*	.35	3.66**
SA of Impact of Other Aircraft on Flight*	.23	2.46*
TLX: Mental Demands of the Task	.21	2.25*

Note. * denotes significant predictor at p<.05
** denotes significant predictor at p<.01

Pilots who had higher self-ratings for situation awareness pertaining to how other aircraft in the system would impact their ability to fly the "perfect" pattern also showed better incursion management. A linear regression model revealed that together, pilot rating and self-rated mental workload and situation awareness accounted for 26% of the variance in runway incursion outcome, $F(3, 87)=10.22$, $p<.001$. Two additional self-rated items of situation awareness showed marginal ($p<.1$) association with incursion management: awareness of task priorities and position of ownship relative to other aircraft. However, these items did not contribute significantly to the regression model.

Discussion

This study investigated whether pilot age, experience and pilot self-rating of situation awareness and workload predicted surprise simulated runway incursion outcome. With respect to age, the oldest group and the youngest pilot group with low ratings achieved mean scores that represent high risk (score less than 6/10) when encountering a runway incursion upon final approach. The marginally significant cubic relationship between age and runway incursion outcome scores supports this conclusion with larger samples perhaps defining this relationship further. Pilot rating contributed the most unique variance in runway incursion management scores. Subjective measures indexing a pilot's self-rating for mental demands and comprehensive situation awareness factors pertaining to the interpretation of the impact of other aircraft on one's own flight were also significant predictors of the outcome of the surprise incursion upon final approach.

Stakeholders addressing runway safety can capitalize on the evidence that strategies to reduce risk for pilots when encountering a runway incursion are best directed towards pilots with lower rating and poorer self-evaluation of situation awareness of other aircraft and poorer appreciation of the mental demands required in high workload conditions. A variety of methods to encourage targeted pilot self-evaluation between regulated biennial assessments of competence should be considered.

Acknowledgements

Funding for this research is provided in part by scholarships to K. Van Benthem from the Social Sciences and Humanities Research Council of Canada and the J. James Mackie Endowment for Graduate Scholarships, Ottawa, Canada. Research infrastructure was supported through a Canadian Foundation for Innovation and Ontario Innovation Trust grant to C. M. Herdman. Authors acknowledge Anne Barr and Andrew Staples of for their simulation engineering expertise.

References

Causse, M., Dehais F., Arexis, M. & Pastor, J. (2011). Cognitive aging and flight performances in general aviation pilots. Aging, Neuropsychology and Cognition, *18*(5) 544-561.

Coffey, E., Herdman, C.M. & Brown, M. (2007). Age-related changes in detecting unexpected air traffic and instrument malfunction. *Proceedings of the 14th International Symposium on Aviation Psychology, Columbus, OH.*

Endsley, M.R. (1988). Situation awareness global assessment technique (SAGAT). *Proceedings of the National Aerospace and Electronics Conference (NAECON)*, 789–795. New York: IEEE.

Endsley, M.R. (2000). Theoretical underpinnings of situation awareness: A critical review. In M.R. Endsley & D.J. Garland (Eds.), *Situation awareness analysis and measurement*. Mahwah, NJ: LEA.

Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Human Mental Workload. P.A.M. Hancock, N. Amsterdam, North-Holland: 139-183.

Honeywell Aerospace. (2009). *Honeywell smart runway and smart landing: Reducing the Risk of runway incursion and excursions*. Retrieved April 15, 2011 from www.honeywellrunwaysafety.com. © 2009 Honeywell International Inc.

Hou, M., Kobierski, R., & Brown, M. (2007). Intelligent adaptive interfaces for the control of multiple UAVs**.** *Journal of Cognitive Engineering and Decision Making, 1*(3), 27-28-362.

Jones, D. (2000). Subjective measures of situation awareness. In M.R. Endsley & D.J. Garland (Eds.), *Situation awareness analysis and measurement*. Mahwah, NJ: LEA.

Kennedy, Q., Taylor, J. L., Reade, G. &Yesavage, J. (2010). Age and expertise effects in aviation decision making and flight control in a flight simulator. *Aviation, Space, and Environmental Medicine, 81*, 489-497.

Taylor, R. M. (1990). Situational Awareness Rating Technique(SART): The development of a tool for aircrew systems design. *AGARD, Situational Awareness in Aerospace Operations 17 p(SEE N 90-28972 23-53).*

Taylor, J., O'Hara, R., Mumenthaler, M., Rosen, A. &Yesavage, J. (2005). Cognitive ability, expertise, and age differences in following age-traffic control instructions. *Psychology of Aging, 20*, 17–33.

Vidulich, M.A. (1988). The cognitive psychology of subjective mental workload. Human Mental Workload. P.A. Hancock and N. Meshkati. Amsterdam, NL, Elsevier Science Publishers B.V. (North-Holland): 219-229.

Vidulich, M. A., Crabtree, M. S., & McCoy, A. L. (1993). Developing subjective and objective metrics of pilot situation awareness. In Proceedings of the 7th International Symposium on Aviation Psychology (pp. 896-900).

ASSESSING THE EFFECTS OF OFF-NOMINAL CONDITIONS
ON NEXTGEN AIR TRAFFIC CONTROL OPERATIONS

Michael W. Sawyer, Katherine A. Berry, & Edward M. Austrian
Fort Hill Group
Washington, DC

The Next Generation Air Transportation System (NextGen) proposes many new tools and capabilities to meet goals of increasing the capacity, safety, and efficiency in the National Airspace System (NAS). This effort aims to assess the potential effects of off-nominal conditions on human performance in the NextGen environment. To complete this assessment, a comprehensive list of off-nominal conditions with potential NextGen consequences was developed. This condition list was then compared against the changes proposed in each NextGen Operational Improvement to determine the potential positive and negative effects on human performance in terms of safety and efficiency. The most frequently cited off-nominal conditions with potential negative effects on human performance included: Incorrect/Missing Information in Data Block/flight plan, Inadvertent Sector Overload, Conflict/Proximity/Other alert activates erroneously, and Runway Closure. These conditions represent key areas that could have crosscutting impact on controller performance and should be utilized to develop requirements for NextGen midterm concepts.

The Federal Aviation Administration (FAA) is conducting a transformation of the National Airspace System. NextGen aims to improve the convenience and dependability of air travel while increasing safety and reducing environmental impact (FAA, 2012). NextGen plans to meet these goals by introducing a variety of new systems and capabilities. The introduction of each new system and capability into the NAS, especially when considering the system-wide impacts of many concurrent development activities, offers the possibility to increase the human contribution to risk in the NAS (Sawyer, Berry, & Blanding, 2011). Research into the effects of these NextGen changes is needed to address the potential for both positive and negative impacts on the safety of the NAS.

When assessing the potential safety impacts associated with changes to the NAS, it is critical to consider the way systems and procedures will interact with the operational conditions under which controllers function (Berry & Sawyer, 2012). Controllers rarely manage traffic under ideal conditions. Therefore safety assessments must consider the various types of conditions under which the system will operate. For example, while a special approach procedure could safely increase capacity under nominal operating conditions, an unanticipated runway closure during those same procedures might increase the safety risk to an unacceptable level. Additionally, the presence of adverse weather conditions coupled with a closed runway might shift the level of risk associated with a special approach procedure to an unacceptable level. Each NextGen OI should consider these types of alternate operating conditions to ensure these NextGen improvements will not lead to unacceptable levels of risk and potential adverse outcomes.

Off-Nominal Conditions

Ideal or nominal conditions refer to the baseline conditions or primary mode of operations for a given system. These conditions are considered notional and often represent a best-case scenario under which a system will perform. A condition being described as nominal does not necessarily mean that it is the most common operating condition. Rather the term nominal represents a baseline or notional set of conditions from which comparisons to other sets of conditions may be made. System designers often design the initial phase of a system to operate under these nominal conditions representing only the best-case conditions for the system. It is important for system designers to expand design to incorporate off-nominal conditions since systems rarely, if ever, operate in a nominal environment. The multitude of NAS systems coupled with environmental conditions and operational needs make it quite quite rare for controllers to operate for extended periods of time in nominal conditions.

As can be seen in Figure 1, any conditions outside of the nominal set of conditions are therefore considered off-nominal conditions. The spectrum of conditions within the off-nominal set is wide ranging from the common, easily manageable conditions (e.g. alternate runway configuration), to the abnormal conditions (e.g. sector exceeds MAP value), to the emergency conditions (e.g. accident on runway) (Burian, 2008). The common off-nominal conditions along with nominal conditions construct what is considered to be normal operations and represent the situations controllers routinely manage without significant adverse effects. Abnormal and emergency conditions are considered non-normal operations and represent situations that occur infrequently, require significant attention, and potentially involve adverse outcomes. The purpose of these categories is to provide a framework for understanding the spectrum of off-nominal conditions, not to draw specific distinctions between whether a condition is considered common or abnormal. Any given condition could potential stretch across all three categories depending on the specific factors and characteristics of a given situation or facility.

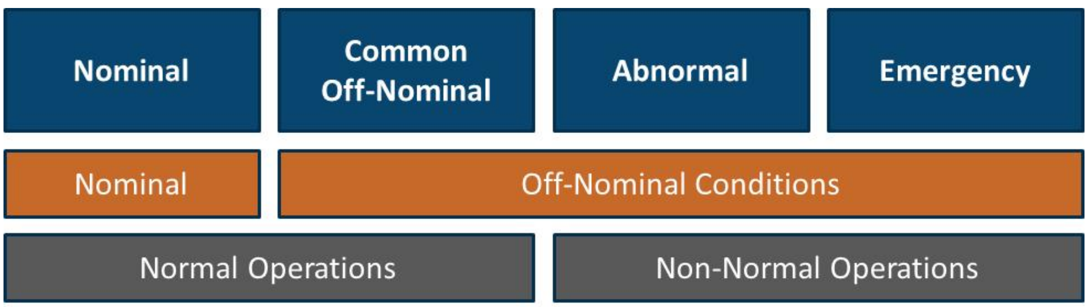


Figure 1. Nominal and Off-Nominal Conditions

Purpose

The purpose of this assessment is to assess the high-level effects of off-nominal conditions on human performance in NextGen operations. This will include identifying new off-nominal conditions present in NextGen operations and assessing the effect of existing off-nominal conditions on controller performance in a NextGen mid-term environment. The results of this analysis should be utilized to create design requirements and identify research needs related to minimizing the adverse effects of off-nominal conditions during NextGen operations.

Development of Off-Nominal Conditions List

In order to assess the potential effects of off-nominal conditions on NextGen operations, it is necessary to first identify the specific conditions to be considered in the assessment. A wide variety of initial off-nominal scenarios and conditions were collected. These initial off-nominal conditions covered the spectrum from common off-nominal conditions, such as mild turbulence, to emergency conditions, such as ATC zero. These initial off-nominal conditions, previous off-nominal research, as well as the experiences and opinions of subject matter experts were combined to create a master list of candidate off-nominal conditions. The master off-nominal conditions list was then consolidated by eliminating duplicate conditions and grouping similar conditions based on the effects of the conditions on controller performance and their potential relevance in proposed NextGen operations.

Off-Nominal Conditions List

The finalized listing of off-nominal conditions contained forty-seven conditions, which were grouped into five broad categories. The first category, Airport/Airspace Conditions, describes the off-nominal conditions present in the operating environment that generally affect any aircraft operating in a given environment. The Automation Performance category contains conditions related to the performance characteristic of the systems and tools utilized in the NAS. This category includes failures, degraded modes, or unexpected performance by various types of automation. The Event – Aircraft Initiated category describes conditions where the actions or performance requirements of an aircraft impact the performance of a controller. These conditions cover the spectrum of conditions from relative common and low-consequence, Aircraft Push Takes Longer Than Expected, to the more urgent Aircraft Declares Emergency. Conditions in the Event – ATC Initiated include changes to the operating environment that are caused by the actions of a controller. This includes conditions such as Aircraft Placed in Holding and Aircraft Go-Around (ATC instructed). The final category, Weather/Environmental Conditions, covers the naturally occurring conditions in the environment that inevitably impact traffic such as Icing or Thunderstorms. The complete listing of off-nominal conditions that was developed and used for this assessment is provided below in Table 2.

NextGen Operational Improvement Assessment

The off-nominal conditions identified in the first phase of this assessment were utilized in assessing the impact of off-nominal conditions on NextGen midterm Operational Improvements (OIs). Human factors and Air Traffic Controller subject matter experts reviewed each midterm OI against the list of off-nominal conditions to assess the potential effects on controller performance. For each OI where the off-nominal conditions were deemed to have a potentially adverse impact on human performance, the condition and potential effects were described. The resulting list provided a crosscutting view of the potential effects of off-nominal conditions on NextGen midterm capabilities.

The comparison of OIs to the Off-Nominal conditions list yielded multiple conditions that could potentially affect each OI. The results of the comparison found eighteen of the off-nominal conditions that would potentially affect more than five different OIs. A summary of the off-nominal conditions deemed to potentially impact the largest number of OIs is provided in Table 1.

Discussion

The results of the OI comparison provided a broad overview of the potential impact of off-nominal conditions on controller performance in the NextGen environment. It further highlights the significance of considering the implications of off-nominal conditions during the design phase of NextGen systems. The most frequently cited conditions and examples of their potential impact are provided below.

Incorrect/Missing Information in Data Block/flight plan/flight object. The most frequently cited off-nominal condition that could potentially affect NextGen operations dealt with the presence of incorrect or missing information in the flight object. As the level of automation in the NextGen environment increases, the importance of automation maintaining an accurate record of each aircraft may also increase. Many of the planned tools relating to improving NAS efficiency rely on optimizing route assignments or traffic sequences. Incorrect information regarding an aircraft's route or performance characteristics could result in the automation yielding a sub-optimal recommendation. Implementing these recommendations could easily result in significant additional controller workload associated with tactically modifying the sequence.

Other OIs propose advanced procedures relying on the accuracy of aircraft information in automation. For example, delegated responsibility for in-trail separation will utilize information about the aircraft to determine the appropriate aircraft pairing. Inaccuracies in aircraft type, equipage levels, or route of flight may potentially lead inadequate pairing of aircraft. This could lead to a loss of efficiency, increased controller workload, or potentially create conflict situations that a controller would need to resolve.

Table 1: Count of Off-Nominal Conditions

Off-Nominal Condition	Count
Incorrect/Missing information in data block/flight plan/flight object	13
Inadvertent Sector Overload	9
Conflict/Proximity/Other alert activates erroneously	8
Runway closure	8
Thunderstorm	7
Radar Surveillance Degraded Mode / Failure	7
Aircraft requests an emergency landings / Aircraft declares emergency	7
Human/Animal/Workers in movement area	7
VFR GA Traffic / Airspace Violator	6
Combined position / sectors	6
Loss of radar contact with an aircraft	6
Runway Incursion	6
Conflict/Proximity/Other alert does not activate in a timely fashion	6
Aircraft misses assigned taxiway exit	6
Aircraft lost/unfamiliar with airport surface	6
Aborted take off	6
Foreign Object Debris on Runway	6
Icing (moderate, severe)	5

Inadvertent Sector Overload. Inadvertent sector overload refers to situations where a sector's traffic level has exceeded the pre-defined acceptable traffic level for that sector. There can be many potential causes of inadvertent sector overload, including many of the other off-nominal conditions listed above. Typically in these circumstances, a controller may enlist the help of their supervisor and the traffic management unit to help tactically manage the traffic in their sector until it returns to a manageable level. The increased capacity and efficiency provided by NextGen tools, such as Point in Space Metering (OI 104120), could make tactically managing a sector more difficult due to aircraft spaced closer together. Other proposed changes, such as automating the handoff process (OI 102114), may further complicate the recovery process by potentially requiring this feature to be inhibited to prevent more aircraft from entering an overloaded sector. Incorporating a requirement to consider the traffic level of the receiving sector into the algorithm that determines whether a sector can receive an automatic handoff could potentially reduce the adverse impact of sector overload in the NextGen environment.

Conflict/Proximity/Other Alert Activates Erroneously. Even with the relatively small number of alerts presently utilized by controllers to ensure NAS safety, the effects of erroneously activated alerts represents one of the FAA ATO's top five hazards in the NAS (Teixeira, 2013). A previous review of potential new alerts and alarms in the NextGen midterm identified fourteen additional alerts that could be provided across the En Route, TRACON, and Tower environments (Berry & Sawyer, 2012). Many of the proposed NextGen OIs present controllers with new alerts aimed at improving NAS safety. OI 102114, Initial Conflict Resolution Advisories, includes references to several potential new alerts that may not only identify potential route conflicts, but may suggest recommended advisory actions to mitigate the conflict. The presence of erroneously activated alerts may require the controller to not only assess whether a true conflict exists, but also determine whether the proposed resolution will adequately resolve the conflict and not create additional flow issues within their sector. If the controller is provided with an accurate alert regarding a conflict, but is given a potentially erroneous resolution it may be more difficult for the controller to resolve the conflict than if no resolution had been provided.

Runway Closure. Runway closures occur primarily for the purposes of planned maintenance activities, such as repairs, resurfacing, or repainting. These activities typically involve a considerable amount of coordination between all airport stakeholders to ensure that all parties are aware of the airport conditions. The increased number of surface management tools as proposed in the NextGen midterm creates an additional set of systems which must be made aware of runway closures. Many OIs, such as Enhanced Surface Traffic Operations (OI 104207) and Enhanced Departure Operations (OI 104208), may present new capabilities related to improving taxi and departure efficiency that will require accurate information regarding the status of runways and taxiways. Inaccurate information could easily lead to the tools providing inaccurate recommendations that may require a controller to tactically manage and resolve any potential issues.

Conclusions

This work provides a set of 47 off-nominal conditions to be used for assessing human performance in air traffic control. It further provides a high-level overview of the results of an assessment of the impacts of these conditions on proposed midterm NextGen capabilities. Eighteen of these conditions were found to impact five or more NextGen OIs. These findings should be incorporated into the development of research and design requirements, as well as testing and human-in-the-loop simulation requirements, in order to ensure capabilities are developed for the range of conditions under which they will be used.

Acknowledgements

We would like to acknowledge the FAA's Human Factors Division (ANG-C1) for funding this project and similar work. Additionally, we would like to acknowledge the air traffic control and human factors subject matter experts who help to assess the operational improvements.

References

Berry, K. & Sawyer, M. (2012). Assessing the Impact of NextGen Trajectory Based Operations on Human Performance. In the *Proceedings of the 4th Annual Applied Human Factors and Ergonomics Conference*, 2012, San Francisco, CA.

Burian, B.K. (2008). Perturbing the system: Emergency and off-nominal situations under NextGen. International Journal of Applied Aviation Studies, 8(1), 114-127.

FAA. (2012). NextGen Implementation Plan. Retrieved March 2012, from http://www.faa.gov/nextgen

Sawyer, M., Berry, K, & Blanding, R. (2011). Assessing the Human Contribution to Risk in NextGen. In the *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2011, Las Vegas, NV.

Teixeira, J. (2013). Improving Aviation Safety: An Air Traffic Control Perspective [PowerPoint Slides]. Retrieved from https://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/ safety/

Table 2. List of Off-Nominal Conditions

Airport / Airspace Conditions	
– Human/Animal/Workers in movement area	– Foreign Object Debris on runway
– Ground stop at destination	– Runway closure
– Taxiway closure	– Excessive workload
– Airport closure	– Combined position/sector
– Runway checks in progress	– Ground delay program in place
Automation Performance	
– Altimeter settings not updated to actual conditions	– Loss of radar contact with an aircraft
– Automation inadequately performs primary function	– GPS system failure/degraded mode
– Blocked ATC communications	– Keyboard, trackball failure / degraded mode
– Alert activates erroneously	– Radar/Surveillance failure / degraded mode
– Alert does not activate in a timely fashion	– ATC Zero
– Incorrect/Missing information in data block/flight object	
Event – Aircraft Initiated	
– Aircraft airborne without release or aircraft misses flow time	– VFR GA Traffic / Airspace Violator
– Aborted take off	– Runway incursion
– Aircraft must return to gate	– Aircraft delays execution of taxi/takeoff
– Aircraft declares emergency	– Aircraft lost/unfamiliar with airport surface
– Aircraft requests priority handling	– Aircraft misses assigned taxiway exit
– Aircraft automation system failure / degraded mode	– Aircraft push takes longer than expected
– Pilot requests a change of destination	– Pilot response to TCAS RA
Event – ATC Initiated	
– Aircraft go-around (ATC instructed)	– Aircraft arriving/departing in opposite direction
– Aircraft placed in holding	– Special approach/departure procedures
– Handoff conducted with aircraft position not in accordance with LOA or SOP	– Aircraft at abnormal altitude for direction of flight
– Aircraft crosses sector boundary on unanticipated trajectory	
Weather / Environment	
– Icing (moderate, severe)	– Turbulence
– Low visibility	– Wind shear
– Thunderstorm	

PILOT COGNITIVE FUNCTIONING AND TRAINING OUTCOMES

Thomas R. Carretta[1] Raymond E. King[2] Paul D. Retzlaff[3]
Erica Barto[4] Malcolm James Ree[4] Mark S. Teachout[5]

The predictive validity of scores from two cognitive functioning tests, the Multidimensional Aptitude Battery and MicroCog, was examined for initial pilot training outcomes. In addition to training completion, academic grades, daily flying grades, check ride grades, and class rank were available for graduates. Mean score comparisons and correlations in samples of between 5,582 and 12,924 trainees across the two tests showed small, but statistically significant, relationships with training performance. The results pointed to general cognitive ability as the main predictor of training performance. Comparisons with results from studies involving US Air Force pilot aptitude tests showed lower validities for these cognitive functioning tests. This finding likely occurred because the pilot aptitude tests measure additional factors (e.g., aviation knowledge/experience, psychomotor) that are predictive of training success, but not measured by these cognitive functioning tests, which were designed primarily to be used for clinical assessment.

Measures of cognitive ability have been a mainstay in military pilot aptitude batteries since WWI (Carretta & Ree, 2003). Although the specific content and administration mode vary, cognitive ability has shown a consistent relation with pilot performance (Carretta & Ree, 2003; Hunter & Burke, 1995; Martinussen, 1996). More recently, Paullin, Katz, Bruskiewicz, Houston, and Damos (2006) conducted a comprehensive review of aviation testing and selection for the US Army that included both cognitive and personality tests. They recommended the US Army follow the lead of the US Navy and US Air Force in their use of selection tests and that they focus on measures of intelligence, cognitive ability, and information processing. Howse and Damos (2011) updated that work with a comprehensive, 275-page annotated bibliography published through the Air Force Personnel Center. These reviews and other studies (Olea & Ree, 1995; Ree & Carretta, 1996; Zierke, 2012) have shown intelligence and cognitive ability to be crucial to pilot training performance. Additional predictors include aviation knowledge/experience, psychomotor ability, and, perhaps, personality (Carretta & Ree, 2003; Hunter & Burke, 1995; Martinussen, 1996).

US Air Force (USAF) Pilot Trainee Selection

All USAF pilot training applicants must pass a rigorous Class I flight physical (USAF, 2011) to be eligible for selection. Medically qualified applicants are evaluated for training suitability on measures of aptitude and officership (Weeks & Zelenski, 1998). USAF Academy (USAFA) cadets are evaluated by faculty and staff, who consider academic, physical, and military performance. Applicants commissioned through the Reserve Officer Training Corps (ROTC) or Officer Training School (OTS) are administered the Air Force Officer Qualifying Test (AFOQT; Drasgow, Nye, Carretta, & Ree, 2010) and Test of Basic Aviation Skills (TBAS; Carretta, 2005). The AFOQT Pilot composite, several TBAS subtest scores, and a measure of flying experience are combined n a regression-weighted equation to create a measure of pilot training aptitude called the Pilot Candidate Selection Method (PCSM; Carretta, 2011). For ROTC, medically qualified applicants are ranked on an Order of Merit score based on the PCSM score, field training, physical fitness, college Grade Point Average (GPA), and commander's ranking. OTS pilot candidate selection uses the "whole person" concept, where applicants receive points for experience/leadership, education/aptitude, and potential/adaptability. All of these selection procedures emphasize high intelligence, whether it involves acceptance into the USAFA, a high GPA, a high AFOQT score, or the impression a candidate makes on a selection board.

Air Force Officer Qualifying Test. The current AFOQT (Form S) has 11 cognitive subtests used to create five composites: Verbal (V), Quantitative (Q), Academic Aptitude (AA), Pilot (P), and Combat Systems Officer (CSO). The V and Q composites are used to qualify civilians and prior-enlisted USAF personnel for officer commissioning through the OTS and ROTC programs. The P and CSO composites are used to qualify applicants who pass other educational, aptitude, and physical requirements for aircrew training. The AFOQT has a hierarchical factor structure and measures general cognitive ability (*g*) and the lower order factors of verbal, math, spatial, aircrew interest/ aptitude, and perceptual speed (Drasgow et al., 2010). It has been validated for officer training (Roberts & Skinner, 1996), aircrew training (Carretta, 2008, 2013; Carretta & Ree, 1995, 2003; Olea & Ree, 1994), and for several non-

[1] Air Force Research Laboratory, Wright-Patterson AFB, OH; [2]USAF School of Aerospace Medicine, Wright-Patterson AFB, OH;; [3]Eagle Applied Sciences, San Antonio, TX; [4]Operational Technologies Corp, San Antonio, TX; [5]University of the Incarnate Word, San Antonio, TX

aircrew jobs (Carretta, 2010). Its predictive validity for pilot training comes from its measurement of cognitive ability and pilot job knowledge (Carretta & Ree, 2003; Olea & Ree, 1994; Ree & Carretta, 1996).

Despite its effectiveness for measuring cognitive ability and its utility for officer and aircrew training qualification, AFOQT scores are not easily interpretable in ideographic assessment. USAF clinicians prefer tests such as the Multidimensional Aptitude Battery (MAB; Jackson, 1984, 1988) to the AFOQT in such assessment due to its similarity to tests such as the Wechsler Adult Intelligence Scale - Revised (WAIS-R). Clinicians find the MAB relatively easy to use to make pre- and post-incident comparisons due to its similarity to the WAIS-R.

Cognitive Testing. While accession procedures focus on intelligence, so does much of the ideographic assessment of pilot candidates. The USAF Medical Flight Screening (MFS) program screens pilot candidates prior to Specialized Undergraduate Pilot Training (SUPT). In addition to ophthalmic and cardiac diagnostic procedures, several cognitive and personality tests are administered (King, Barto, Ree, & Teachout, 2011; King, Barto, Ree, Teachout, & Retzlaff, 2011). The primary purpose of the cognitive tests is to archive cognitive functioning data for future use. The intent is to develop a registry against which future testing might be compared. The psychological portion of MFS includes both traditional measures of intelligence and computerized cognitive tasks.

As the primary purpose of the psychological testing is to enable potential ideographic assessment, there has been little emphasis on training outcomes. To date, the MFS cognitive tests have not been validated against pilot training outcomes. Boyd, Patterson, and Thompson (2005), however, evaluated some of the tests against aircraft type later flown. Usually, fighter/bomber aircraft advanced training assignments are offered to those highest in class rank in primary jet training. Class rank accounts for much of the variance in advanced training assignments although other factors (e.g., number of fighter/non-fighter training slots, student preferences, and Guard/Reserve pilots flying what their squadrons fly) also affect advanced training assignments. Boyd et al. (2005) compared one of the MFS cognitive tests, the MAB (Jackson, 1998), and one of the personality tests, the NEO PI-R (Costa & McCrae, 1985), to airframe assignment (fighter, bomber, and airlift/tanker). Small (Cohen, 1988), but statistically significant differences were observed between the groups, with the mean IQs for fighter pilots 2 to 3 points (about .13 to .20 SDs) higher than for airlift/tanker pilots. Using the means, SDs, and sample sizes reported by Boyd et al., we converted the differences to a correlation statistic (Lipsey & Wilson, 2000). The mean difference in verbal, performance, and full scale IQ between those assigned to fighters vs. airlift/tankers were equivalent to correlations of .14, .15, and .18. These results suggest intelligence has a modest relationship with advanced training assignment.

Purpose

The purpose of the current study was to determine the extent to which two tests used to assess cognitive functioning and typically the domain of clinical assessment, the MAB and MicroCog, predict USAF initial (T-6) pilot training outcomes. Separate validation studies were done in order to maximize the sample sizes for each test.

Methods

Participants

Participants were USAF personnel selected for SUPT that had tested on the MAB, MicroCog, or both. The sample sizes were 12,924 for the MAB and 5,582 for the MicroCog. All participants were college graduates or near completion of college. Sample demographics were similar for the two studies. Of those reporting demographic data, 91% were male. They had a mean age of 23 years, and 99% were 30 years of age or less. Eighty-four percent reported they were white. Test administration either occurred at the USAFA or at the USAF School of Aerospace Medicine (USAFSAM) prior to entry into SUPT. The T-6 completion rate was about 89.5% for both samples.

Measures

Multidimensional Aptitude Battery. The MAB (Jackson, 1984, 1998) is a broad-based test of cognitive ability patterned after the WAIS–R (Wechsler, 1981). The full-scale IQ scores for the MAB and WAS-R are highly correlated ($r = .91$; Conoley & Kramer, 1989; Jackson, 1984). The MAB can be individually or group administered and requires less than 1.5 hours. Its 10 subtests produce three scores: full-scale IQ (FSIQ), verbal IQ (VIQ), and performance IQ (PIQ). The IQ scores have a mean of 100 and a SD of 15 in the general population. Test-retest reliability for the IQ scores ranges from .94 to .98 (Jackson, 1998) for a retest interval averaging 45 days. The FSIQ score has been shown to measure *g* in several age groups (Wallbrown, Carmin, & Bartlett, 1988, 1989). Chappelle, McDonald, Thompson, McMillan, and Marley (2010) examined the MAB for USAF gunship sensor operators and found no mean differences between training graduates and eliminees.

MicroCog. The MicroCog (Powell, Kaplan, Whitla, Weintraub, Caitlin, & Funkenstein, 1993) is a computer-administered cognitive functioning test that assesses a range of cognitive behaviors such as reaction time and memory. The primarily purpose of the test was to assess clinical pathology in patients. While the MAB is a classic IQ test, the MicroCog comes more from a clinical neuropsychological perspective (Vanderploeg, 2000).

The MicroCog has 18 subtests combined to create nine index scores. The indices take two forms, domain-based and higher-order summary scores. The five domains are Attention/Mental Control, Memory, Reasoning/Calculation, Spatial Processing, and Reaction Time. The four higher-order summary scores are Information Processing Speed (IPS), Information Processing Accuracy (IPA), General Cognitive Functioning (GCF), and General Cognitive Proficiency (GCP). IPS and IPA reflect a potential two-factor structure of the subtests. GCF and GCP are purported to represent general cognitive ability, where GCF is a function of the two Information Processing scores and GCP is a summation of the Proficiency scores of all the subtests (Powell et al., 1993). The Information Processing and General Cognitive indices generally correlate with the WAIS-R in the .50s.

Chappelle, Ree, Barto, Teachout, and Thompson (2010) compared the MAB and MicroCog using structural equation models. They concluded that both tests have a factor representing g. The MicroCog only produced one factor, suggesting there is less specificity to the scores than may be desired by clinicians or researchers. Inasmuch as the MicroCog appears to measure only one factor, and due to space limitations, we focused on the four higher-order summary scores in our analyses.

Pilot training criteria. Several SUPT initial jet training (T-6) performance criteria were examined. There were three dichotomous training completion scores for graduates and eliminees: graduation/elimination, graduation/flying training deficiency (FTD) elimination, and graduation/drop on request (DOR) elimination. Several additional criteria were available only for graduates: academic grades, daily flying grades, check flight grades, and class rank. Class rank is a weighted average of academic, daily flying, and check flight grades. In computing class rank, flying grades get more weight than academic grades and check flight grades get more weight than do daily flying grades.

Analyses

Analyses began with examination of the means and SDs for the cognitive test scores. Univariate statistics were used to determine the relations of the cognitive test scores to the training performance criteria. All statistical analyses used a .05 Type I error rate and one-tailed tests. Next, the observed correlations between the test scores and training criteria were corrected for range restriction using the multivariate method (Lawley, 1943). The MAB and MicroCog scores could not be corrected to the same reference group as the participants lacked a common selection test (e.g., AFOQT). As a result, the data for each test were corrected to the respective normative group. After correction for range restriction, the correlations involving the test scores and training completion criteria were corrected for dichotomization (Cohen, 1983). The correlations involving the test scores and pilot training grades were corrected for unreliability (Hunter & Schmidt, 2004) of the training criteria ($r_c = \frac{r_{xy}}{\sqrt{r_{yy}}}$). The reliability of the training grades was estimated at .80 based on results from similar studies that examined academic grades (Kuncel, Hazlett, & Ones, 2001, 2004). The correlations corrected for range restriction and reliability of the training grades provide a theoretical estimate of the predictiveness of the test scores when a perfectly reliable criterion is available.

Results

Study 1: Multidimensional Aptitude Battery

Graduation vs. elimination. The overall graduation rate was 89.6% (11,579/12,924). When only graduates and either FTD or DOR eliminees were included the graduation rates were 95.4% (11,579/12,138) and 95.9% (11,579/12,079) respectively.

The MAB IQ scores were severely range restricted compared to the normative values where the means and SDs are 100 and 15. The IQ scores for graduates and each of the eliminee groups were high at about 120 (about 1.33 SDs above the normative mean) and the variances of the scores were much less than the normative values. For the FSIQ score the variance for the trainees was about 18% of the normative value. All mean score differences between graduates and eliminees favored graduates, but were small (i.e., about 2 points for the IQ scores). Despite this, all mean score comparisons were statistically significant. Larger differences occurred for graduates vs. FTD eliminees than for graduates vs. DOR eliminees. This result may be because DOR elimination may occur for reasons not related to ability (e.g., motivation). Examination of the observed correlations indicated all effect sizes were small (< .10; Cohen, 1989). While very large samples ensure sufficient statistical power, very small differences will be statistically significant yet may offer little practical predictive power. Low point-biserial correlations for the IQ scores reinforce the small mean score differences. A .083 correlation was observed between the FSIQ score and the graduation/elimination criterion. It should be noted that the training eliminees included medical and self-elimination losses, so the group distinctions in this analysis are not as clear as desired.

Table 1 summarizes the observed and corrected correlations. As expected, the correlations increased in magnitude after correction. For example, the correlation between the FSIQ score and graduation/elimination was .083 in the observed data, increased to .192 after correction for range restriction, and to .323 after correction for both range restriction and dichotomization of the criterion. Similar trends occurred for the other MAB scores and criteria.

Table 1. *MAB: Observed and Corrected Correlations with Training Completion*

Score	Graduation/All Eliminees			Graduation/FTD Eliminees			Graduation/DOR Eliminees		
	r	r_c	r_{fc}	r	r_c	r_{fc}	r	r_c	r_{fc}
FSIQ	.083[b]	.208	.295	.072[b]	.158	.264	.039[b]	.106	.202
VIQ	.057[b]	.185	.262	.042[b]	.115	.192	.022[a]	.101	.192
PIQ	.079[b]	.199	.282	.074[b]	.165	.275	.042[b]	.104	.198

Notes. The column headings indicate observed correlations (r), correlations corrected for range restriction (r_c), and correlations corrected for both range restriction and dichotomization of the criterion (r_{fc}). Statistical significance was tested only for the observed correlations. [a]$p < .05$; [b]$p < .01$

Training grades. As shown in Table 2, all observed correlations between the test scores and training grades were statistically significant. FSIQ had the strongest observed correlation for all of the training grades. The strongest correlations for the IQ scores occurred for academic grades (e.g., FSIQ; r = .233). The FSIQ correlation with class rank, which is a weighted average of the academic and flying training grades, was .157. All correlations increased in magnitude after correction for range restriction and again after correction for both range restriction and reliability of the criteria. After correction for both range restriction and reliability of the criteria, FSIQ was correlated .551 with academic grades, .316 with daily flying grades, .282 with check flight grades, and .374 with class rank.

Table 2. *MAB: Observed and Corrected Correlations with Training Grades*

Score	Academic Grades			Daily Flying Grades			Check Flight Grades			Class Rank		
	r	r_c	r_{fc}	r	r_c	r_{fc}	r	r_c	r_{fc}	r	r_c	r_{fc}
FSIQ	.233[b]	.505	.564	.124[b]	.257	.287	.110[b]	.235	.262	.157[b]	.325	.363
VIQ	.224[b]	.497	.555	.084[b]	.229	.256	.083[b]	.217	.242	.123[b]	.308	.344
PIQ	.164[b]	.428	.478	.120[b]	.252	.281	.099[b]	.222	.248	.138[b]	.304	.339

Notes. The column headings indicate observed correlations (r), correlations corrected for range restriction (r_c), and correlations corrected for both range restriction and reliability of the criterion (r_{fc}). Statistical significance was tested only for the observed correlations. N = 11,579; [a]$p < .05$; [b]$p < .01$

Study 2: MicroCog

Graduation vs. elimination. The overall graduation rate was 89.4% (4,992/5,582). When only graduates and either FTD or DOR eliminees were included the graduation rates were 93.5% (4,992/5,238) and 96.1% (4,992/5,194) respectively.

Although the MicroCog scores were affected by range restriction, the amount of restriction was less than that for the MAB. Both tests have means and SDs of 100 and 15 in their respective normative samples. However, whereas the average means and SDs for the MAB IQ scores were about 120 and 6.4 for pilot trainees, the average means and SDs for the MicroCog scores were about 104 and 11. The variances of the MAB and MicroCog scores were respectively about 18% and 54% that for their respective normative populations. The difference in amount of restriction on the two tests was likely due to differences in the composition of the normative groups. MicroCog population norms are based on scores corrected for age and education level. All mean score comparisons between graduates and eliminees favored graduates and were statistically significant for the analyses involving all eliminees

Table 3. *MicroCog: Observed and Corrected Correlations with Training Completion*

Score	Graduation/All Eliminees			Graduation/FTD Eliminees			Graduation/DOR Eliminees		
	r	r_c	r_{fc}	r	r_c	r_{fc}	r	r_c	r_{fc}
IPS	.068[b]	.125	.201	.077[b]	.129	.252	.024[a]	.050	.112
IPA	.053[b]	.137	.220	.063[b]	.147	.287	.015	.050	.112
GCF	.083[b]	.168	.270	.097[b]	.178	.347	.026[a]	.063	.141
GCP	.091[b]	.165	.266	.102[b]	.171	.334	.031[a]	.065	.145

Notes. The column headings indicate observed correlations (r), correlations corrected for range restriction (r_c), and correlations corrected for both range restriction and dichotomization of the criterion (r_{fc}). Statistical significance was tested only for the observed correlations. [a]$p < .05$; [b]$p < .01$

and FTD eliminees. Smaller mean score differences occurred between graduates and DOR eliminees. As with the MAB, all point-biserial correlations effect sizes were small (< .10; Cohen, 1989).

Table 3 summarizes the observed and corrected correlations. Even after correction for both range restriction and dichotomization, only eight of 12 correlations were above .20; only two were above .30.

Training grades. As with the MAB, all observed correlations between the test scores and training criteria were statistically significant. See Table 4. GCF and GCP demonstrated the highest predictive validities averaged across the training criteria. The strongest correlations for three of the four MicroCog scores occurred for academic grades. After correction for both range restriction and reliability of the criteria, GCF was correlated .341 with academic grades, .285 with daily flying grades, .251 with check flight grades, and .333 with class rank.

Table 4. *MicroCog: Observed and Corrected Correlations with Training Grades*

Score	Academic Grades			Daily Flying Grades			Check Flight Grades			Class Rank		
	r	r_c	r_{fc}	r	r_c	r_{fc}	r	r_c	r_{fc}	r	r_c	r_{fc}
IPS	$.075^b$	.164	.183	$.137^b$	.208	.232	$.102^b$	.164	.183	$.139^b$	.219	.244
IPA	$.220^b$	.314	.351	$.093^b$	.191	.213	$.097^b$	.182	.203	$.138^b$	.246	.275
GCF	$.206^b$	.305	.341	$.165^b$	.255	.285	$.149^b$	.225	.251	$.202^b$	.298	.333
GCP	$.204^b$	.299	.334	$.170^b$	.249	.278	$.148^b$	.219	.244	$.201^b$	.289	.323

Notes. The column headings indicate observed correlations (r), correlations corrected for range restriction (r_c), and correlations corrected for both range restriction and reliability of the criterion (r_{fc}). Statistical significance was tested only for the observed correlations. N = 4,992; $^a p < .05$; $^b p < .01$

Discussion

This study examined the relations between two cognitive functioning tests and pilot training outcome. Overall, the results were consistent with prior studies of the relations between cognitive ability and pilot training outcomes (e.g., Carretta & Ree, 2003; Hunter & Burke, 1995; Martinussen, 1996; Zierke, 2012). The MAB and MicroCog both assess *g* (Chappelle et al., 2010), as does the AFOQT. The lower validities for the MAB and MicroCog compared to USAF pilot aptitude tests (i.e., AFOQT and PCSM) was likely due to additional factors measured by the pilot aptitude tests. A joint confirmatory factor analysis of the AFOQT and MAB revealed that each test had a hierarchical structure (Carretta, Retzlaff, Callister, & King, 1998). The higher-order factor in the AFOQT has been identified as general cognitive ability (*g*) (Drasgow et al., 2010). The correlation between the higher-order factors from the two tests was .98 indicating that both measured *g*. Although both tests measure *g*, and include verbal, spatial, and perceptual speed content, the AFOQT also includes tests of aviation knowledge not found in the MAB (Carretta et al., 1998). It is likely that the MicroCog does not assess such unique factors either (Chappelle et al., 2010) . The higher validities for the AFOQT Pilot and PCSM composites compared with the MAB and MicroCog is likely due to their measurement of additional factors shown to be related to pilot training performance (e.g., aviation knowledge/experience and psychomotor) that are not included in the MAB and MicroCog, which are primarily designed for clinical assessment. Aviation knowledge and experience may be an indirect measure of motivation.

Nevertheless, the MAB and MicroCog scores demonstrated predictive validity against most of the training criteria. For example, after correction for both range restriction and reliability of the criteria, the MAB FSIQ score was correlated .564 with academic grades, .287 with daily flying grades, .262 with check flight grades, and .363 with class rank. The MicroCog also showed generally significant results when compared against training criteria, but had lower validities than the MAB after correction.

Neither the MAB nor the MicroCog, however, was an effective predictor of graduation versus DOR elimination. This result was expected as DOR elimination is affected by both ability and motivation and neither of the tests assesses motivation. It should be noted that these cognitive functioning tests were not administered with the primary purpose of predicting training outcomes. Rather, their purpose was to baseline cognitive functioning for potential future ideographic comparisons.

As with other occupations (Schmidt & Hunter, 1998), pilot training performance is affected by both ability (can do) and motivation (will do) factors. Cognitive aptitude tests measure the "can do" component of achievement, while factors such as prior aviation experience and specialized job-related knowledge sought by the applicant, and personality, measure the "will do" component. As no USAFSAM cognitive functioning tests directly assess aviation motivation, a future study will examine personality and its incremental validity in the prediction of flying training performance when used in combination with measures of cognitive ability.

References

Boyd, J. E., Patterson, J. C., & Thompson, B. T. (2005). Psychological test profiles of USAF pilots before training vs. type aircraft flown.

Aviation, Space, and Environmental Medicine, 76, 463-468.
Carretta, T. R. (2005). *Development and validation of the Test of Basic Aviation Skills (TBAS),* AFRL-HE-WP-TR-2005-0172. Air Force Research Laboratory, Human Effectiveness Directorate, Wright-Patterson AFB, OH.
Carretta, T. R. (2008). *Predictive validity of the Air Force Officer Qualifying Test for USAF air battle manager training performance*, AFRL-RH-WP-TR-2009-0007. Air Force Research Laboratory, Human Effectiveness Directorate, Wright-Patterson AFB, OH.
Carretta, T. R. (2010). Predictive validity of the Air Force Officer Qualifying Test for non-rated officer specialties. *Military Psychology, 22,* 450-64.
Carretta, T. R. (2011). Pilot Candidate Selection Method: Still an effective predictor of US Air Force pilot training performance. *Aviation Psychology and Applied Human Factors, 1,* 3-8.
Carretta, T. R. (2013). Predictive validity of pilot selection instruments for remotely piloted aircraft training outcome. *Aviation, Space, and Environmental Medicine, 84,* 1-7.
Carretta, T. R., & Ree, M. J. (1995). Air Force Officer Qualifying Test validity for predicting pilot training performance. *Journal of Business and Psychology, 9,* 379-88.
Carretta, T. R., & Ree, M. J. (2003). Pilot selection methods. in P. S. Tsang & M. A. Vidulich, eds., *Human factors in Transportation: Principle s and Practice of Aviation Psychology,* pp. 357-396. Mahwah, NJ: Erlbaum.
Carretta, T. R., Retzlaff, P. D., Callister, J. D., & King, R. E. (1998). A comparison of two US Air Force pilot aptitude tests. *Aviation, Space, and Environmental Medicine, 69,* 931-935.
Chappelle, W., McDonald, K., Thompson, W., McMillan, K., & Marley, M. (2010). *Multiple Aptitude Battery-II normative intelligence test data that distinguish U.S. Air Force AC-130 gunship sensor operators,* AFRL-SA-BR-TR-2010-0006. U.S. Air Force School of Aerospace Medicine, Brooks City-Base, TX.
Chappelle, W., Ree, M. J., Barto, E. L., Teachout, M. S., & Thompson, W. T. (2010). *Joint use of the MAB-II and MicroCog for improvements in the clinical and neuropsychological screening and aeromedical waiver process of rated USAF pilots,* AFRL-SA-BR-TR-2010-0002. US Air Force School of Aerospace Medicine, Brooks City-Base, TX.
Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7,* 249-253.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
Conoley, J. C., & Kramer, J. J. (1989) *The tenth mental measurements yearbook*. Lincoln, NE: The University of Nebraska Press.
Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory Manual.* Psychological Assessment Resources, Odessa, FL.
Drasgow, F., Nye, C. D., Carretta, T. R., & Ree, M. J. (2010). Factor structure of the Air Force Officer Qualification Test Form S: Analysis and comparison with previous forms. *Military Psychology, 22,* 68–85.
Howse, W. R., & Damos, D. L. (2011). *A bibliographic database for the history of pilot training selection*, AFCAPS-FR-2011-0010. Air Force Personnel Center, Randolph AFB, TX.
Hunter, D. R., & Burke, E. F. (1995). *Handbook of pilot selection.* Brookfield, VT: Avebury Aviation.
Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis.* Thousand Oaks, CA: Sage.
Jackson, D. N. (1984). *Multidimensional Aptitude Battery: Manual*. Research Psychologists Press, Port Huron, MI.
Jackson, D. N., (1998). *Multidimensional Aptitude Battery-II: Manual.* SIGMA Assessment Systems, Port Huron, MI.
King, R. E., Barto, E., Ree, M. J., & Teachout, M. S. (2011). *Compilation of pilot personality norms*, AFRL-SA-WP-TR-2011-0008. U.S. Air Force School of Aerospace Medicine, Wright-Patterson AFB, OH.
King, R. E., Barto, E., Ree, M. J., Teachout, M. S., & Retzlaff, P. D. (2011). *Compilation of pilot cognitive ability norms,* AFRL-SA-WP-TR-2012-0001. U.S. Air Force School of Aerospace Medicine, Aeromedical Research Department, Wright-Patterson Air Force Base, OH.
Kuncel, N. R., Hazlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127,* 162-181.
Kuncel, N. R., Hazlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86,* 148-161.
Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, 62, (Section A, Part 1),* 28-30.
Lipsey, M. W., & Wilson, D. B. (2000). *Practical Meta-Analysis*. Sage, Thousand Oaks, CA.
Martinussen, M. (1996). Psychological measures as predictors of pilot performance: A meta-analysis. *International Journal of Aviation Psychology, 6,* 1-20.
Paullin, C., Katz, L., Bruskiewicz, K. T., Houston, J., & Damos, D. (2006). *Review of aviator selection*, Technical Report 1183. U.S. Army Research Institute for the Behavioral and Social Sciences, Arlington, VA.
Olea, M. M., & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than *g*. *Journal of Applied Psychology, 79,* 845-51.
Powell, D. H., Kaplan, E. F., Whitla, D., Weintraub, S., Caitlin, R., & Funkenstein, H. H. (1993). *MicroCog: Assessment of cognitive functioning (Version 2.1) Manual*. Psychological Corporation, San Antonio, TX.
Ree, M. J., & Carretta, T. R. (1996). Central role of *g* in military pilot selection. *International Journal of Aviation Psychology, 6,* 111-123.
Roberts, H. E., & Skinner, J. (1996). Gender and racial equity of the Air Force Officer Qualifying Test in officer training school selection decisions. *Military Psychology, 8,* 95-113.
Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262-274.
U.S. Air Force (2011). *Medical examinations and standards,* Air Force Instruction 48-123. Washington, DC: Department of the Air Force.
Vanderploeg, R. D., ed. (2000). *Clinician's guide to neuropsychological assessment,* 2nd ed. Mahwah, NJ: Erlbaum.
Wallbrown, F. H., Carmin, C. N., & Barnett, R. W. (1988). Investigating the construct validity of the Multidimensional Aptitude Battery. *Psychological Reports*, *62*, 871-878.
Wallbrown, F. H., Carmin, C. N., & Barnett, R. W. (1989). A further note on the construct validity of the Multidimensional Aptitude Battery. *Journal of Clinical Psychology, 45,* 429-433.
Wechsler, D. (1981). *Wechsler Adult Intelligence Scales – Revised (WAIS-R).* The Psychological Corporation: NY.
Weeks, J. L., & Zelenski, W. E. (1998). *Entry to USAF undergraduate flying training,* AFRL-HE-AZ-TR-1998-0077. Brooks AFB, TX: Air Force Research Laboratory, Training Effectiveness Branch, Warfighter Training Research Division.
Zierke, O. (2012). *Predictive validity in aviation psychology: Really not much more than g?* Paper presented at the 30th Meeting of the European Association of Aviation Psychology. Villasimius, Sardinia, Italy.

MEASURING DISTRIBUTION OF ATTENTION AS A PART OF SITUATIONAL AWARENESS - A DIFFERENT APPROACH

Meierfrankenfeld, Katrin
German Air Force Institute of Aviation Medicine
Fürstenfeldbruck, Germany
Greß, Werner
German Air Force Institute of Aviation Medicine
Fürstenfeldbruck, Germany
Vorbach, Tina
German Air Force Institute of Aviation Medicine
Fürstenfeldbruck, Germany

This paper outlines a different approach to measure pilot aptitudes during flight simulator missions. An algorithm was developed to assess a candidate's distribution of attention beyond observation technique, eye-tracking or multi-dimensional tracking (e.g. altitude, speed, heading), thus getting rid of typical measurement problems. The algorithm used to evaluate candidate's distribution of attention in Phase III, German Armed Forces' third phase of aircrew selection consisting of simulator flights in a typical training scenario, is a mere time measure. The following article describes its construction as well as advantages and disadvantages.

Situation awareness (SA) is a key concept in aviation psychology. Crashes are frequently explained by loss of SA (e.g. Endsley & Garland, 2000; Nullmeyer, Stella, Montijo, & Harden, 2005; Jones, D.G., & Endsley, M.R., 1996). Aircraft interface upgrades are justified by assumed increases in SA (Vidulich, 2003), and a lot of research deals with SA. Nevertheless, it is yet unclear what SA in fact is. A widely accepted definition is proposed by Endsley (2000): [SA 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." (p. 5). Distribution of attention might be seen as first level of SA. It can also be called attentional flexibility or allocation of attention, which is often measured by visual scanning behavior (Bellenkes, Wickens & Kramer, 1997). Salmon, Stanton, Walker, and Green (2006) state: [Measurement of eye movements is] "recording the process that operators use in order to develop SA" (p.234). For use in personnel selection, it is inconvenient to wear eye trackers because of weight, costs, and usual drop-outs due to technical problems. Furthermore, "to see" does not necessarily mean "to perceive". Another approach to measure distribution of attention is multi-dimensional tracking (e.g. altitude, speed, heading) based on deviation measures, leading to the problem of units: How to build a score of e.g. 10° heading deviation, 5 knots speed deviation and 100 feet altitude deviation? In flight training and selection programs, expert ratings are frequently used (e.g. FAA flight test standards, FAA, 2002). As typical observer mistakes can occur, objective data might be helpful to ensure objectiveness and reliability. This paper will describe the development of an approach to measure distribution of attention objectively.

Measuring Distribution of Attention – a different approach

This section gives a brief overview over German Armed Forces' aircrew selection process, leading to a description of the simulator flight investigated here and steps towards measuring distribution of attention.

German Armed Forces' Aircrew Selection

German Armed Forces' (GAF) aircrew selection procedure consists of three phases. Phase I and II include the assessment of basic aptitudes and the aviation-medical examination. Phase III (fixed wing) is more complex. It consists of one week simulator-based screening in a typical training scenario: Candidates prove their skills both in 4 simulator-flight missions with increasing workload and in academic training. As in real flight training, a briefing, a demonstration and a practice phase and subsequent debriefings prepare candidates for their check phases. The aim is to evaluate aptitudes, to propose specific cockpit assignments (e.g. jet pilot, weapon system operator/ navigator, transport pilot), and to minimize attrition rate during basic flight training. The aircrew selection process works quite well, as long term evaluation shows: Attrition rates during flying training are very low (e.g. in ENJJPT: 2007 to 2012: 5,4% total and 3,8% due to flying deficiencies). Per year, approximately 200 applicants are tested at Phase III fixed wing.

Flight Simulator used in this study: The FPS/F

The FPS/F (Aviation Psychological Pilot Selection System/ Fixed Wing) is a flight simulator consisting of 4 cockpits with lockable canopies, a spherical projection dome with 200° horizontal and 45° vertical field of view, a 5-channel high resolution projection system, a multi-functional display with all basic flight instruments plus a master caution panel for malfunctions and a radio panel (Figure 1). The instructor's consoles enable monitoring the applicant's activities and performance. Video protocols as well as mission logs are used for debriefing purposes. Data can be analyzed at an evaluation station.

First steps on the way to measure distribution of attention objectively

First results from Mission 2 (Figure 2) are reported. Mission 2 consists of traffic patterns with full stop landing. As the required X-check varies with the demands of the tasks, sets of variables necessary for a proper X-check for every maneuver in Mission 2 were defined. Acceptable deviations around ideal values were defined in a second step. Examples are shown in Table 1. Third, an algorithm was defined that computes the proportion of time where distribution of attention fails. This means the candidate violates one or more of the defined ranges in heading, altitude, speed, vertical speed and/or angle of bank **and** is **not** correcting. Afterwards, composite scores were calculated. These composite scores for distribution of attention reflect distribution of attention during the whole mission. As performance during Mission 2 changes due to concentration, practice effects or difficulty, composite scores for each pattern were calculated, too. In spite of grouping by time (pattern 1, 2 or 3), we also grouped by maneuver type. This means composite scores were calculated for turns, legs and other maneuvers.

Figure 1. The flight simulator used in Phase III/ fixed wing (FPS/F) consists of cockpits with a high quality screen comprising the field of view (200° horizontal, 45° vertical) (left), and the multifunctional display showing expanded instrumentation as well as touchscreen and radio (right).

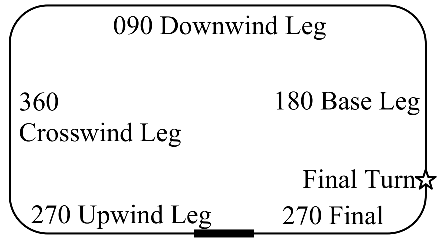


Figure 2. Mission 2 consists of 3 pattern flights starting on runway 27. In pattern 3, a gear emergency occurs on downwind leg and on base, new R/T is required.

Expert ratings in Phase III. Experts grade each maneuver, each pattern, the whole Mission 2, as well as distribution of attention on scales from 1, "excellent", to 7, "unsatisfactory". Further aptitudes assessed during Phase III are not reported in this article. Success in Phase III as reported here is based on performance and progress from Mission 1 to Mission 4. It ranges from 1, "excellent", to 7, "unsatisfactory", with grades 6 and 7 indicating no proposal for any cockpit position.

Hypothesis. An objective time-based score for distribution of attention should correlate with the experts' ratings of distribution of attention, with performance in Mission 2 and with overall success at Phase III. The last mentioned correlation is expected to be only small to medium, as Mission 2 is only one of four missions plus theoretical tests that lead to the decision if someone is proposed to become a GAF's crew member.

Results

Data from all candidates who passed Phase III from January to April 2012 were used. From 52 applicants, 1 passed with "A/excellent", 4 with „B/good", 11 with „C/average", 8 with „D/marginal", and 28 with "U/unsatisfactory". There was 1 female candidate. Applicants were young adults, aged from 18 to 24 (*M* = 20, *SD* = 1,96). 83% have passed A-levels, 7 had a secondary school level certificate (13%) and 2 (4%) had an advanced technical college entrance qualification.

Experts graded distribution of attention ranged from 2 to 7 (*M* = 4.8, *SD* = 1.42). Performance in Mission 2 ranged from 1 to 7, with a mean of 4.9 (*SD* = 1.57). Average grade in pattern 1 was 4,8 (*SD* = 1.31), 4,7 in pattern 2 (*SD* = 1.32) and 5,1 in pattern 3 (*SD* = 1.6). The score for distribution of attention in Mission 2 could theoretically range between 0 (all relevant parameters within the acceptable ranges **or** correcting towards the desired values) and 57 (always exceeding the acceptable ranges **and** not correcting). In this sample, the score ranged from 9.85 to 31.56 (*M* = 20.6, *SD* = 4.59), with lower scores indicating a better distribution of attention. The mean pattern-wise scores are 7.60 for pattern 1 (*SD* = 2.16); 7.30 for pattern 2 (*SD* = 1.77) and 7.44 for pattern 3 (*SD* = 2.06).

Correlations between reported scores are shown in Table 2. The computed score and the experts' ratings correlate .70, which is significant with $\alpha < .01$ and large (Cohen, 1988): The smaller the score that is the better the distribution of attention, the better the expert's rating of distribution of attention. Furthermore, the computed composite score and success in Phase III correlates .53 ($\alpha < .01$): As expected, a low composite score (meaning good distribution of attention) and good grades (indicating good performance in Phase III) are associated. Anyway, the expert's grades for distribution of attention correlate .72 with results in Phase III ($\alpha < .01$).

Correlations between expert's ratings of distribution of attention and pattern-wise and maneuver-wise composite scores were medium to high and all significant ($\alpha < .05$; see Table 2). Among those correlations, the correlation with the composite score of pattern 2 and with the composite score consisting of legs in pattern 3 are the most high (> .60, $\alpha < .01$).

Conclusion and Discussion

First results concerning the development of a time-based measurement of distribution of attention are reported. Correlations between the computed scores and expert ratings were – as expected – large (Cohen, 1988) and significant. Thus, it seems to be a promising approach. Yet, some questions remain. At first, the scores presented here are mere composite scores without any weighting. Weighting maneuvers depending on their difficulty would be reasonable: Turns are more dynamic than straight and level legs, thus a more fluent and quick distribution of attention might be needed. Legs in pattern 3 are more difficult than in pattern 1 and 2 because of the occurrence of an emergency and unexpected R/T, thus indicating distribution of attention although distractors are calling for attention. Correlations (Tab. 2) might point in that direction: Pattern 2 and 3 and legs in pattern 3 seem to be most important. Further investigation is needed to test these hypotheses. Second, the candidate's progress in distribution of attention might be interesting. As can be seen in the results section, applicants perform best at pattern 2 and the least at pattern 3. Does distribution of attention differ during patterns, too? And how is its progress during whole Phase III, from Mission 1 to Mission 4? The third issue to be discussed is about limits and chances of performance measures. The score is performance-based – time of (non-) performance is measured. In flight training, performance is main criteria, too. But performance measures are always contaminated: Psychomotor skill, decision making, speed of information processing, concentration, speed of automation, stress level, aggressiveness and other aptitudes influence performance in such a complex scenario as Mission 2, too. This problem cannot be solved in this article; for discussion see e.g. Pew (2000). Anyway, relationships between aptitudes assessed in Phase III and the composite score should be analyzed, giving hints concerning the score's construct validity. Furthermore, the relationship between expert ratings and composite scores should be examined in more detail: While objective measures are supposed to be more reliable than subjective measures because the latter might be contaminated by the human observer, objective measures reduce information, thus they might be failing to explain complex decision processes and/ or scenarios (e.g. Bell & Lyon, 2001). How is the computed score working, does it explain variance in applicant's performance beyond observer's ratings, is there a gain in incremental validity? Is the higher correlation between success in Phase III and experts' ratings of distribution of attention due to implicit weighting, to methodology effects – observer's scales are similar, and both rely on observation method- , is observation biased, or is its predictive validity as to later training results in fact higher? A promising approach for further improvements of the selection process.

Table 1.

Sets of Criteria (Examples) and Acceptable Deviations Around Them as Basis for the Distribution of Attention Score.

Maneuver	Set of variables	Range
Level Flight	Altitude 1000 feet	*+/- 20 feet*
	Indicated Air Speed 130 kts	*+/- 3 kts*
	Required HDG	*+/- 1°*
Level Turn	Altitude 1000 feet	*+/- 20 feet*
	Indicated Air Speed 130 kts	*+/- 3 kts*
	AOB 30°	*+/- 2°*
	R/O: Required HDG	*+/- 1°*

Note. N = 52. HDG = Heading. AOB = Angle of Bank. R/O = Roll out.

Table 2.

Correlations between Computed Scores of Distribution of Attention, Expert Ratings and Success at Phase III

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 Expert Rating	1														
2 Score Total	*.70***	*1*													
3 Score Pattern 1	*.50***	*.78***	*1*												
4 Score Pattern 2	*.63***	*.85***	*.51***	*1*											
5 Score Pattern 3	*.58***	*.81***	*.37***	*.62***	*1*										
6 Score Turns 1	*.41***	*.66***	*.80***	*.40***	*.37***	*1*									
7 Score Turns 2	*.51***	*.68***	*.46***	*.62***	*.58***	*.44***	*1*								
8 Score Turns 3	*.43***	*.66***	*.26*	*.54***	*.8***	*.32**	*.54***	*1*							
9 Score Legs 1	*.43***	*.65***	*.71***	*.36***	*.47***	*.63***	*.54***	*.37***	*1*						
10 Score Legs 2	*.53***	*.69***	*.42***	*.78***	*.51***	*.36***	*.40***	*.44***	*.28**	*1*					
11 Score Legs 3	*.61***	*.81***	*.50***	*.59***	*.89***	*.39***	*.57***	*.61***	*.6***	*.52***	*1*				
12 Score Dyn 1	*.31**	*.52***	*.76***	*.39***	*.10*	*.34**	*.16*	*.00*	*.18*	*.31**	*.22*	*1*			
13 Score Dyn 2	*.35**	*.53***	*.27*	*.76***	*.32**	*.14*	*.07*	*.25*	*.04*	*.40***	*.27*	*.36**	*1*		
14 Score Dyn 3	*.35**	*.74***	*.14*	*.36***	*.73***	*.17*	*.27*	*.29**	*.16*	*.27*	*.53***	*.02*	*.25*	*1*	
15 Success_PhaseIII	*.72***	*.53***	*.30**	*.49***	*.51***	*.23*	*.33**	*.45***	*.21*	*.39***	*.40***	*.23*	*.35**	*.4***	*1*

Notes. N = 52. Pearson product-moment correlation coefficients are reported. The computed score for distribution of attention at Mission 2 is called Score Total. Correlations between scores for legs, turns, and other maneuvers (dyn = dynamic) as well as pattern-wise scores are also reported. Expert Rating refers to expert rating of distribution of attention. * $p \leq .05$ ** $p \leq .001$.

References

Bell, H.H., & Lyon, D.R. (2001). Using observer ratings to assess situation awareness. In Mica R. Endsley, Danlie J. Garland (Eds.). Situation Awareness Analysis and Measurement (pp. 129-146). Mahwah, NJ: Lawrence Erlbaum Associates.

Bellenkes, A.W., Wickens, C.D., & Kramer, A.F. (1997). Visual scanning and pilot experience: The role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine*, *68*, 569-579.

Cohen, J.W. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn). Hillsdale, NJ: Lawrence Erlbaum Associates.

Endsley, M.R. (2000). Theoretical Underpinnings of Situation Awareness: A critical Review. In Mica R. Endsley, Danlie J. Garland (Eds.). *Situation Awareness Analysis and Measurement* (pp. 3-32). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M.R., & Garland, D.G. (2000). Pilot situation awareness training in general aviation. In *Proceedings of the 14th triennial congress of the International Ergonomics Association and the 44th annual meeting of the Human Factors and Ergonomics Society* (pp. 2/357-2/360). Santa Monica, CA: Human Factors and Ergonomics Society.

Federal Aviation Administration. (2002). *Private pilot practical test standards for airplane* (FAA Pub. No. FAA-S-8081-14A). Retrieved from http://www.faa.gov/training_testing/testing/airmen/test_standards/

Jones, D.G., & Endsley, M.R. (1996), Sources of Situation Awareness Errors in *Aviation. Aviation, space and environmental Medicine,* 67 (6), 507-512.

Nullmeyer, R.T., Stella, D., Montijo, G.A., & Harden, S. W. (2005). Human Factors in Air Force flight mishaps: Implications for change. In *Proceedings of the 27th annual Interservice/Industry Training, Simulation, and Education Conference* (Paper No. 2260). Airlington, VA: National Training Systems Association.

Pew, R.W. (2001). The state of situation awareness measurement: Heading toward the next century. In Mica R. Endsley, Danlie J. Garland (Eds.). Situation Awareness Analysis and Measurement (pp.33-47). Mahwah, NJ: Lawrence Erlbaum Associates.

Salmon, P.M., Stanton, N.A., Walker, .H., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics*, *37*(2), 225-238.

Vidulich, M.A. (2003). Mental Workload and Situation Awareness. In Pamela S. Tsang, Michael Vidulich (Eds.). Principles and Practice of Aviation Psychology (pp. 115-146). Mahwah, NJ: Lawrence Erlbaum Associates.

Detecting Fatigue in Commercial Flight Operations using Physiological Measures

Lisa C Thomas
Kimberly Craig
Christopher Gast
Robert Grube
Mike Muhm
Emma Romig
Boeing Commercial Airplanes
Seattle, WA

The purpose of this study is to determine whether any technology exists to unobtrusively, reliably, and accurately detect symptoms of fatigue in real time before fatigue affects performance. Airline pilots are fitted with a variety of physiological measurement devices (e.g. EEG, blink rate, etc.) that have been demonstrated in the literature to be related to fatigue. Each crew of two pilots performs simulated gate-to-gate flight operations under rested and fatigued conditions, during which physiologic and performance parameters are continuously monitored. In addition, audio, video, and simulator data are recorded for post-session evaluation. Ultimately, if one or more technologies proves effective, we can incorporate it into the flight deck for real-time fatigue detection capability as part of a larger fatigue risk management system. The usefulness of this type of approach extends beyond the commercial flight deck to any work environment that requires multi-shift or other non-traditional scheduling.

"In the operational environment, sleepiness-related performance deficits are typically the result of interactions involving multiple factors such as recent and long-term sleep-wake history, the circadian rhythm of alertness, effects of time on task, cognitive workload, and individual differences in resilience and sensitivity to sleep loss. Such factors should be taken into consideration when devising operational work-rest schedules and when assessing individual[s] who complain of deficits in alertness and performance" (Balkin, 2011).

For any fatigue risk management program to be effective, it is important to gain better understanding of all of the various factors that contribute to fatigue and, more importantly, when the effects of fatigue may negatively affect the pilots' abilities to cope with those situations. A tired pilot may not always make mistakes and in fact may rarely make mistakes, but he is at greater risk in all situations than an alert pilot. If it is possible to quantify how each factor contributes to the experience of fatigue, insight may be gained into how best to mitigate those factors and/or the resulting fatigue.

We are particularly interested in how the differences in the types of commercial flight operations may affect fatigue in different ways. Long-haul pilots must deal with circadian disturbances with time zone transitions, which result in disrupted sleep schedules and ineffective recovery periods (Bourgeois-Bougrine et al, 2003). Even when they have opportunities to sleep on the aircraft during longer flights, it is not always restful given the noise, temperature, lighting, and other factors which reduce the quality of sleep (Rosekind, Miller, Gregory, & Dinges, 2000). For short-haul pilots, relatively high workload contributes to fatigue (Bourgeois-Bougrine et al, 2003); "because takeoffs and landings are extremely task-intensive, it logically follows that a flightcrew member who has performed six sets of takeoffs and landings will be more fatigued than the flightcrew member who has performed only one takeoff and landing" (FAA, 2011). If fatigue has different causal factors and different manifestations, our ability to measure and even predict the effects for various types of operations may aid the design of more effective mitigation strategies.

Several studies have examined the physiological symptoms as well as performance of fatigued and rested pilots during real operations (Thomas et al, 2006; Samel, Wegmann, & Vejvoda, 1997; Wright & McGown, 2001; Signal, Gale, & Gander, 2005). Thomas and his colleagues collected data from crews who provided sleep/wake diary, actigraphy, psychomotor vigilance task (PVT) performance, and subjective ratings during four international flights. They found that the amount of sleep obtained by pilots in the 24 hours preceding the flight had significant effects on both self-reported fatigue levels as well as PVT performance; specifically, that less sleep resulted in higher fatigue ratings and slower response times in the PVT task.

A review by Mathis & Hess (2009) recommends that researchers interested in measuring fatigue and sleepiness create "a battery of subjective and objective tests to answer a specific question in order to achieve the most appropriate description," since each test measures a particular aspect of a particular factor and any one test run in isolation may misrepresent (or miss) what is happening physically, physiologically, or cognitively on a larger scale.

Physiological and performance measures of fatigue and sleepiness

There are several strategies for detecting fatigue and/or sleepiness, and determining fatigue effects on performance of tasks of varying cognitive or physical effort. Objective physiological measures of physiological responses related to fatigue include brain activity, eye movements and eye-closure related measures, heart rate related measures, actigraphy, postural measures, and voice/speech analysis. Some of these measures are more intrusive than others (devices that require electrodes or are worn on the body), which limit their usefulness in an applied setting. A number of companies are marketing devices that are far less obtrusive to track and detect fatigue in situations where fatigue effects could be potentially harmful, such as driving, long-haul trucking, shipping, and aviation, with varying levels of success and acceptance by the targeted user community.

Objective performance measures of fatigue effects include performance on PVT and other stimulus-response tasks, tasks along the cognitive effort continuum (e.g. simple tasks such as basic addition or word games through complex problem solving), pattern-matching and monitoring tasks, and decision-making. In some cases, the performance measured is the subject's primary task (e.g. flight path deviations made by a fatigued pilot) and as such is operationally relevant; however it can be difficult to determine when fatigue is having an effect if the task does not lend itself to purely objective standards (e.g. was the landing really well done, good, or just acceptable) or there are other factors contributing to a change in performance (e.g. is a pilot's delay in setting the flaps due to fatigue or a competing task). In other cases, the tasks are somewhat arbitrary and are exclusively used as a fatigue detection metric and serve no operational purpose (e.g. the PVT), but they tend to have clearer, more explicitly defined parameters of good/fatigued performance.

Subjective measures of fatigue effects include a variety of self-report rating scales and sleep diaries, where people indicate the frequency and duration of their sleep periods and assign ratings to the quality of sleep obtained. Subjective measures are likely to be less reliable compared to objective measures, since they are more heavily influenced by those factors that mask fatigue or sleepiness, as noted in studies that show low subjective ratings of fatigue (subject reports feeling sufficiently alert) but clear fatigue effects in objective measures (subject's PVT scores are slower than their typical "alert" scores).

Current Study

The current study is intended to evaluate a variety of physiological measurement devices and other sensors embedded into a flight deck and used in the context of commercial flight deck operations. Data from these tools and from flight performance will be analyzed to determine whether one or more of these devices can be used to effectively and reliably detect fatigue, and possibly even predict performance decrements due to fatigue effects.

We have selected a wide range of embedded physiological devices that have been validated by prior research to effectively detect symptoms of fatigue and/or sleep deprivation. Data from these devices will be collected and evaluated to determine whether fatigue effects and symptoms can be effectively detected in the context of long and short haul flight operations. We will also be collecting performance data based on pilot interaction with the flight deck, flight performance as evaluated using a Flight Operations Quality Assessment (FOQA) analysis, and decision-making as rated by a panel of subject matter experts who observe the pilots and rate their actions and decisions against a set of criteria or expectations. FOQA algorithms and criteria can be applied to evaluate flight performance (e.g. deviations from the planned flight path, quality of landings, etc).

We will compare all our physiological measures against subjective ratings (KSS, SP) and objective (PVT) measures that are well-established measures of fatigue. In addition, we will compare the subjective/objective measures of fatigue with the measures of performance.

Expected Results

- We expect to see an association between pilot errors and physiological evidence of fatigue effects (e.g. long blinks, head-nodding, low scores on PVT, KSS, SP, EEG evidence of micro-sleep events, etc).
    - The time series of physiological measures will be compared to the time series of pilot performance measures (incorporating adjustment for relative fatigue state) to determine what predictive ability the physiological measures may hold. Both cumulative and instantaneous measurements of each variable will be examined for their ability to predict pilot error with high selectivity.
- We expect to see associations between evidence of fatigue and/or sleepiness in the various embedded devices. For example, if the EEG shows a change in brain activity concordant with increased drowsiness or sleepiness, there will be a concurrent increase in percent eye closure and in posture modification during that time period. Similarly if the EOG shows slow eye movement, the eyelid position tracking data should indicate long fixations or slow blink rates.
    - The time series of each physiological measurement will be compared to one another to determine the level of agreement. For EEG measurements, the PSG method will be considered the standard to which other EEG measurements are compared.
- We expect to see evidence of deteriorations in crew resource management (CRM) as quantified by LOSA-style evaluation during time periods when fatigue symptoms are detected by the various devices.
- We expect to observe pilot error rates that are dependent on fatigue state (rested/fatigued) and cumulative workload (early/late) and their interaction.

Method

Subjects. Thirty two crews consisting of a Captain and First Officer from the same airline company will be recruited for inclusion in the study. In addition, up to 10 additional pilots will be recruited to allow final improvements to elements of the study. All participants must be qualified in

Boeing 777 aircraft. Because of the paucity of female commercial airline pilots, the selection of participants will be limited to males. Additional demographic information, such as age, flight hours, and fatigue-relevant habits (e.g., smoking, normal caffeine intake, known sleep abnormalities) will be collected.

Apparati. The study will be conducted in the 777 Cab, a high fidelity flight simulator. The 777 Cab is equipped to record data from pilot interactions with the controls and displays and this data will be time-synched to the physiological device data outputs. Physiological measures include EEG (electroencephalograph), EOG (electrooculograph), ECG (electrocardiograph), actigraphy, blink rate, head posture, body posture, and voice samples. Additional measures include psychomotor vigilance test (PVT) scores, subjective fatigue and sleepiness scores, and behavioral observations via audio/video recordings.

Scenarios. Half of the total participants will be assigned to long-haul flight scenarios, and the other half to a series of four shorter flights that will be conducted in succession to simulate short-haul operations. The sessions will be matched in terms of total flight duty time (eight hours), and involve gate to gate operations, from simulated dispatch interactions, through the flight itself, and ending with a taxi to the arrival gate. Further, at the end of the long-haul flights and in the last flight of each short-haul series, the crew will encounter either a medical emergency or a holding pattern scenario.

All crews will be scheduled to arrive at the lab following a minimum of 72 hours off-duty. This is expected to reduce any accumulated fatigue that may have been acquired during their preceding flight schedules and produce a set of participants who are all within a narrow range of restedness relative to their own baseline at the start of the simulator-based data collection phase. Within each group (long-haul and short-haul), pilots will fly two sessions, one each under Rested and Fatigued conditions. The Rested session will be scheduled when the pilots' predicted alertness is at its peak (i.e. normal daytime schedule after a sufficient rest opportunity). The Fatigued sessions will be scheduled when predicted alertness is at a minimum (i.e. overnight). Prior to the Fatigued sessions, the crew will be given a normal sleep opportunity with a required wakeup time, then remain awake for an extended period of time and participate in scheduled activities during the day to eliminate rest opportunities. For half of the crews, the Rested session will be first, followed by the Fatigued session; the other half of crews will experience the Fatigued session first, followed by the Rested (see Table 1 for full breakdown of conditions).

Table 1.
Outline of each unique session type by flight condition, relative fatigue level, and order of fatigue condition.

Crew	**Session 1**	**Session 2**
Crew A	Rested; Long Haul	Fatigued; Long Haul
Crew B	Fatigued; Long Haul	Rested; Long Haul
Crew C	Rested: Short Haul	Fatigued; Short Haul
Crew D	Fatigued; Short Haul	Rested: Short Haul

Procedure. Prior to participation in the simulator-based data collection sessions, the pilots will be asked to wear an actigraph watch and use the Jeppesen CrewAlert app to enter subjective fatigue ratings and perform a psychomotor vigilance task (PVT) periodically throughout each day for 2 weeks.

On the first day of the simulator-based sessions, both pilots will be brought to the flight simulator. They will be outfitted with the selected instruments and run through any calibration steps that are required. Once the devices are in place and working, the pilots will be briefed as though they were going to fly an actual flight. They will interact with (simulated) dispatch, tower personnel, cabin crew, and ground crew while preparing the plane for takeoff, and perform all duties associated with entering and executing a flight plan. During the gate-to-gate operations, they will encounter a number of glitches, anomalies, and other decision points, which are designed to gather information about the pilots' awareness, vigilance, and alertness as they perform these tasks.

At four times during each session, pilots will be asked to perform extra tasks targeted at acquiring subjective and objective data on fatigue and workload. These Fatigue Assessment Breaks consist of performing the PVT, responding to the KSS and SP subjective ratings scales, reading sentences aloud from a provided sheet, looking for changes in event related potential, and three minutes of rest with eyes closed.

They will be served a standard meal (sandwich, chips, cookie, and water) at approximately the half-way point in the session, and will be allowed to leave the cab when necessary for short periods of time. They will also be offered coffee, water, and granola bars periodically throughout the sessions. No formal napping periods will be allowed, however if one or both pilots falls asleep during the session, the experimenters will not intervene.

After the flight session is concluded, the pilots will be returned to their hotel and will be asked to follow the predetermined procedures for either rest or extended wakefulness. The second session will proceed just like the first, with the crew performing the other of the two flight scenarios under similarly challenging conditions.

Results and Discussion

Data collection on this study began in August 2012 and will continue through early 2014. At the time of this paper's acceptance for publication, we have not acquired sufficient data for preliminary analyses. However, initial results will be reported at the time of presentation.

Conclusions

The results of this study are expected to contribute to our understanding of fatigue in the commercial flight deck; how different factors may differentially affect pilots' physiological symptoms and flight operations performance. With a more thorough understanding of what fatigue looks like in the flight deck, we may be able to more effectively design fatigue risk management programs for a variety of risk-inducing situations.

References

Balkin, TJ. (2011). Performance deficits during sleep loss: Effects of time awake, time of day, and time on task. In MH Kryger, T Roth, & WC Demont (Eds.), *Principles and Practice of Sleep Medicine, 5th edition* (Chapter 65). St Louis, MO: Saunders, Elsevier.

Basner, M, & Dinges, DF. (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep, 34*(5), 581-591.

Bourgeois-Bougrine, S, Cabon, P, Gounelle, C, Mollard, R, & Coblentz, A. (2003). Perceived fatigue for short- and long-haul flights: A survey of 739 airline pilots. *Aviation, Space, & Environmental Medicine, 74*, 1072–7.

FAA (2011a). FAA and Industry are taking action to address pilot fatigue, but more information on pilot commuting is needed. *Office of Inspector General Audit Report No. AV-2011-176*, September 2011.

FAA (2011b). Flightcrew member duty and rest requirements. *US DOT FAA Docket No. FAA-2009-1093; Amendment Nos. 117-1, 119-16, 121-357,* December 2011.

FAA Fact Sheet (2010). *Pilot Fatigue.* Retrieved from http://www.faa.gov/news/fact_sheets/news_story.cfm?newsId=11857

Fogt, DL, Cooke, WH, Kalns, JE, & Michael, DJ. (2011). Linear mixed-effects modeling of the relationship between heart rate variability and fatigue arising from sleep deprivation. *Aviation, Space, & Environmental Medicine, 82*, 1104-9.

Hursh, SR, Raslear, TG, Kaye, AS, & Fanzone, JF. (2006). Validation and calibration of a fatigue assessment tool for railroad work schedules, summary report. Washington DC: Federal Railroad Administration.

Hursh, SR, & Van Dongen, HPA. (2011). Fatigue and performance modeling. In MH Kryger, T Roth, & WC Demont (Eds.), *Principles and Practice of Sleep Medicine, 5th edition* (Chapter 66). St Louis, MO: Saunders, Elsevier.

Kaida K, Akerstedt, T, Kecklund, G, Nilsson, JP, & Axelsson, J. (2007). Use of subjective and physiological indicators of sleepiness to predict performance during a vigilance task. *Industrial Health*, 45, 520-526.

Reifman, J, & Gander, P. (2004). Commentary on the three-process model of alertness and broader modeling issues. *Aviation and Space Environmental Medicine, 75*(3, Supplemental), A84-88.

Samel, A, Wegmann, HM, & Vejvoda, M. (1997). Aircrew fatigue in long-haul operations. *Accident Analysis and Prevention, 29*(4), 439-452.

Thomas, MJW, Petrilli, RM, Lamond, N, Dawson, D, & Roach, GD. (2006). Australian long haul fatigue study. *Proceedings of the 59th Annual International Air Safety Seminar (IASS).*

Van Dongen, HPA, & Hursh, SR. (2011). Fatigue, performance, errors, and accidents. In MH Kryger, T Roth, & WC Demont (Eds.), *Principles and Practice of Sleep Medicine, 5th edition* (Chapter 67). St Louis, MO: Saunders, Elsevier.

Van Dongen, HPA, Mott, CG, Huang, JK, et al. (2007). Optimization of biomathematical model predictions for cognitive performance impairment in individuals: accounting for unknown traits and uncertain states in homeostatic and circadian processes. *Sleep, 30,* 1129-1143.

Vogel, AP, Fletcher, J, & Maruff, P. (2010). Acoustic analysis of the effects of sustained wakefulness on speech. *Journal of the Acoustical Society of America, 128*(6), 3747-56.

EFFECTS OF EVENT RATE ON CEREBRAL BLOOD FLOW VELOCITY DURING VIGILANCE PERFORMANCE

Catherine Neubauer[1], Michael B. Dillard[2], Joel S. Warm[3, 4], Gregory J. Funke[4], Matthew Funke[5], Gerry Matthews[1], Samantha L. Epling[6], & Allen W. Dukes[4]

University of Cincinnati[1]
Cincinnati, Ohio

University of Dayton Research Institute[3]
Dayton, Ohio

Naval Medical Research Unit – Dayton[5]
Wright-Patterson Air Force Base, OH

National Research Council[2]
Washington, DC

Air Force Research Laboratory[4]
Wright-Patterson Air Force Base, OH

Ohio Wesleyan University[6]
Delaware, OH

Transcranial Doppler sonography (TCD) was used to assess the effects of event rate on cerebral blood flow velocity (CBFV). Fast (30 events/minute) and slow (5 events/minute) event rates were employed in a 40-minute vigilance task that simulated the control of remotely piloted aircraft. As is typical in vigilance tasks, signal detection declined over time but there was no performance difference in conjunction with the two event rate conditions. Nevertheless, CBFV was greater in the fast than in the slow event rate condition and declined significantly with time on task. These outcomes support previous findings of a close tie between CBFV and the vigilance decrement and suggest the possibility that in regard to event rate, CBFV is a more sensitive index of task demand than performance itself.

Vigilance, or sustained attention tasks require observers to maintain their focus of attention and detect the appearance of critical signals over prolonged periods of time. This aspect of performance is of interest to the Air Force because of the critical role that operator vigilance plays in enemy surveillance, cockpit monitoring, air-traffic control, and airport and border security (Warm, Parasuraman, & Matthews, 2008). A key finding in vigilance performance is that sustained attention is fragile, it wanes over time. This is reflected in what is known as the vigilance decrement, a decline in the speed and accuracy of signal detections with time on task (Warm et al., 2008). The resource model proposed by Davies and Parasuraman (1982) is a major conceptual framework for understanding the decrement. According to that view, the need to make continuous signal/noise discriminations depletes information-processing assets or reservoirs of energy that cannot be replenished in the time available. Hence, the temporal decline in performance efficiency. Support for the resource model comes from findings that vigilance tasks impose a high level of perceived mental workload on observers which increases with increments in psychophysical demand, that vigilance tasks promote high levels of stress, and that the decrement is correlated with observers' feelings of mental exhaustion (Warm et al., 2008; Warm, Finomore, Vidulich, & Funke, in press).

It is important to note that the resource model has been criticized on the grounds that resources are not measured directly. Instead, they are inferred from performance and therefore, explaining performance changes in terms of resource loss represents circular reasoning (Navon, 1984). Recent research has sought to counter this criticism by assessing resources independently of performance using a non-invasive neuroimaging procedure known as Transcranial Doppler sonography (TCD). The TCD procedure employs ultrasound signals to monitor cerebral blood flow velocity (CBFV) or hemovelocity in the middle cerebral arteries which carry about 80% of the blood within each hemisphere (Toole, 1984). When a particular area of the brain becomes active, as in the performance of mental tasks, by-products of this activity, such as carbon dioxide, increase which lead to an increase in CBFV to the region to remove the waste product (Aaslid, 1986). Consequently, TCD offers the possibility of measuring changes in metabolic activity during task performance and it has been used to do that in a wide variety of cognitive, perceptual, and motor tasks (Tripp & Warm, 2007). As reviewed by Warm and his associates (Shaw, Finomore, Warm, & Matthews, 2011; Warm et al., 2008), studies with regard to vigilance have shown that the level of CBFV varies directly with task demand as determined by variations in several task parameters (e.g., memory load, stimulus uncertainty, signal presence/absence), that the vigilance decrement is accompanied by a temporal decline in CBFV, and that these effects occur primarily in the right hemisphere, indicating a right hemispheric system in the functional control of vigilance. Of critical importance are the findings that the temporal decline in CBFV occurs with both visual and auditory tasks indicating that the effect is general in nature (Shaw et al., 2009) and that it only appears when observers are actively engaged in a vigilance task. Bloodflow velocity remains constant over time when participants observe the vigilance display in the absence of a work imperative, revealing CBFV changes to be task related (Warm et al., 2008). The present study was designed to extend the investigation of CBFV and vigilance by examining hemovelocity effects in the context of a critical psychophysical parameter in vigilance - the background event rate - that has not as yet been examined in regard to bloodflow velocity changes.

Vigilance tasks often employ dynamic displays in which the critical signals for detection are embedded within a matrix of recurring neutral background events. An example would be the need to detect the occasional appearance of a slightly longer line in a cascade of short lines. While the background events may be neutral in the sense that they do not require an overt response from the observer, they are not neutral in their effects on signal detection. The frequency of the background events or the background event rate is a key element in determining performance efficiency. The accuracy of signal detections varies inversely with event rate, the vigilance decrement tends to be more pronounced in the context of a fast as compared to a slow event rate, and the background event rate has been shown to be a moderating variable in regard to the effects of other factors such as signal amplitude and the demands of multi-tasking (Davies & Parasuraman, 1982; Warm et al., in press). Findings such as these have led to the view that event rate is probably the prepotent psychophysical factor in regard to vigilance performance (Parasuraman, Warm, & Dember, 1987).

From the framework of the resource model, the degrading effects of increments in event rate can be interpreted in terms of the greater consummation of information processing assets in a fast as compared to a slow event rate condition brought about by the higher frequency of signal/noise discriminations demanded in the fast event rate condition. Given the greater resource demand in a fast as compared to a slow event rate condition, it might be anticipated that the overall level of CBFV and the temporal decline in CBFV would be greater in the context of a fast than a slow event rate. Moreover, given the evidence pointing to a right-hemispheric system in the control of vigilance, it might also be anticipated that the CBFV effects associated with event rate would be lateralized to the right cerebral hemisphere. The present study was designed to test these possibilities.

Method

Participants

Twenty individuals from the Dayton, OH area (10 males and 10 females) served as observers for a single payment of $30. They ranged in age from 18 to 30 years old. All observers had normal or corrected-to-normal vision and were right handed, as measured by the Edinburgh Handedness Inventory (Oldfield, 1971). The experiment was conducted under conditions approved by the WPAFB Institutional Review Board.

Experimental Design

This study employed a 2 (Event Rate) × 2 (Cerebral Hemisphere) × 4 (Periods of Watch) mixed-design. Ten observers were assigned at random to either a slow or fast event rate condition with the restriction that sex was equated within conditions. All observers participated in a 40-minute vigil divided into 4 continuous 10-minute periods of watch.

Vigilance Task

Observers assumed the role of controllers monitoring the flight paths of remotely piloted aircraft (RPA) on a 17-inch visual display terminal (VDT). The vigilance display, which was modeled after that used by Hitchcock et al. (2003), is shown in Figure 1. It consisted of a “city” signified by a solid red circle (10.5 mm in diameter; transluminance = 21.4 cd/m^2) surrounded by a thin white border (0.75 mm thick × 12 mm in diameter), three concentric white outer markers (0.75 mm thick, 28, 53, and 83 mm in diameter, respectively; transluminance = 79.9 cd/m^2), and two lines representing RPA flight paths (1 × 25 mm; transluminance = 30.6 cd/m^2). The displays were presented on a light gray background (transluminance = 29.5 cd/m^2). The Michaelson Contrast Ratio ([maximum luminance – minimum luminance / maximum luminance + minimum luminance]; Coren, Ward, & Enns, 1999) of the RPA flight paths to the background was 1.83% (light gray targets on a light gray background). Each aircraft approached from opposite headings, in either a NW to SE or SW to NE direction. Safe flight paths or neutral events were those in which the flight headings of two RPAs were slightly displaced to the left or right of the center of city so that they would pass each other without collision. Critical signals for detection were RPAs that were aligned on a collision path over the center of the city. Observers were instructed to press the spacebar on a computer keyboard in the presence of a critical signal and to make no response to non-critical signals. Neutral events and critical signals are displayed in the left and right of Figure 1, respectively.

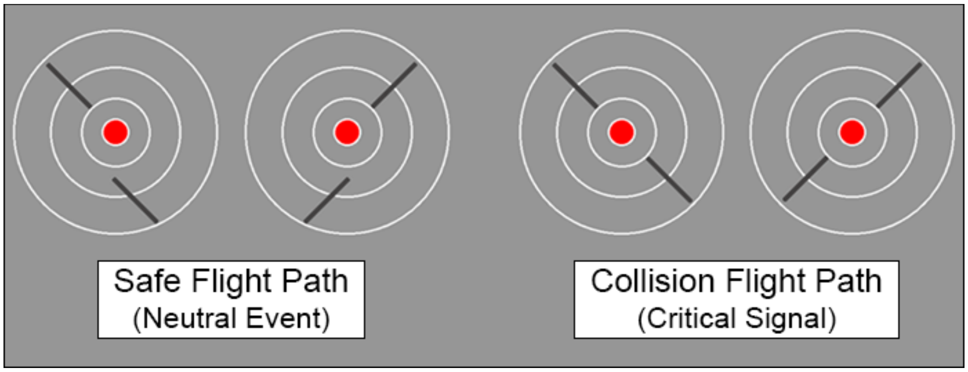


Figure 1. Examples of neutral and critical signals presented in the display. The contrast of the RPA flight paths to the background has been increased in the figure for clarity of presentation.

In the slow event rate condition the display was updated 5 times per minute while in the fast event rate condition it was updated 30 times per minute. In both conditions, the dwell time per stimulus event was 80 ms. Responses made within 1200 ms of the onset of a critical signal were considered as correct detections or “hits.” All other responses were considered as errors of commission or “false alarms.” Ten critical signals were presented per period in all conditions. Each experimental session was preceded by a 5-minute resting period during which CBFV was recorded. CBFV was assessed bilaterally using TCD. As recommend by Aaslid (1986), baseline scores for CBFV were determined by averaging the last 60 seconds of the resting period. Following the resting period, observers in all conditions received a 5-minute practice session. During practice, feedback was provided by a computer generated female voice, which indicated correct detections, false alarms, and misses. Observers were required to detect at least 5 out of 10 critical signals during this time in order to be included in the study. Audio feedback was removed during the main portion of the experimental session. Observers surrendered all timepieces and electronic devices upon reporting for the experiment. Stimulus presentations and vigilance response recording in all conditions were controlled by a Dell PC running Windows XP.

Experimentation was conducted in a 2.48 × 2.45 × 2.16 m windowless sound attenuated booth. The VDT was mounted on a table 70 cm directly in front of the seated observer (visual angles for the VDT and the stimulus display were 32.87° and 6.79°, respectively) Ambient illumination in the testing booth was 2.5 cd/m^2, provided by two 17-watt fluorescent bulbs, occluded on all sides and positioned above and adjacent to the seated observer to minimize glare on the VDT.

Hemovelocity Measurement

A Nicolet Companion III TCD machine was used to index changes in CBFV. CBFV was measured bilaterally using two 2-mHz ultrasound transducers. Measurements were taken from the left and right medial cerebral arteries (MCAs).The transducers were secured in a plastic bracket which was attached to an adjustable plastic headband and secured to the observers head, located dorsal and immediately proximal to the zygomatic arch along the temporal bone. Aquasonic 100-brand gel was placed between the transducer and the observer’s skin to amplify the ultrasound signal. In the current study, the MCA was generally monitored at depths of 50-55 mm. CBFV measures were recorded by the TCD unit at approximately 1 Hz.

Results

Performance Efficiency

Mean numbers of correct detections and false alarms in the two event rate conditions are plotted as a function of periods of watch in Figures 2 and 3, respectively. It is evident in both figures that the scores for the event rate conditions were similar to each other and that they declined over time. Separate 2 (Event Rate) × 4 (Periods) split-plot analyses of variance (ANOVAs) of the correct detection and false alarm data revealed that in both cases the main effect for periods of watch was statistically significant, $F_{Correct\ Detections}$ (2.69, 48.44) = 5.91, $p < .01$, $\eta^2 = .25$; $F_{False\ Alarms}$ (2.81, 50.62) = 4.08, $p < .05$, $\eta^2 = .19$, while the main effect for event rate and the event rate × periods interaction were not significant, $p > .05$ in each analysis. In these and the subsequent ANOVA of the bloodflow data, the Box correction was employed when needed to adjust for violations of the sphericity assumption (Field, 2009).

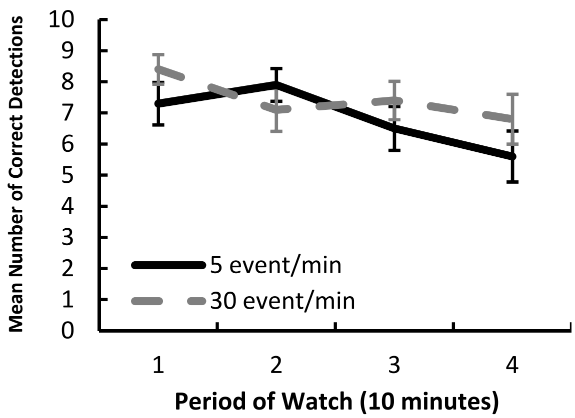


Figure 2. Mean number of correct detections for the slow and fast event rate conditions as a function of periods of watch. Error bars are standard errors.

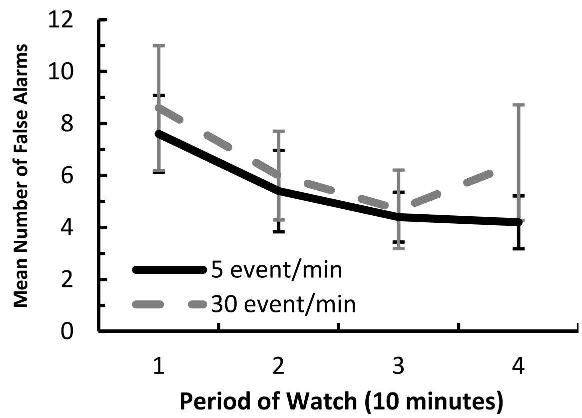


Figure 3. Mean number of false alarms for the slow and fast event rate conditions as a function of periods of watch. Error bars are standard errors.

Hemodynamics

CBFV was measured within the left and right hemispheres for all combinations of event rate and periods of watch. In order to control for individual variability, CBFV was expressed as a proportion of the last 60 seconds of the 5-minute resting baseline (Aaslid, 1986). Mean CBFV scores for the two event rate conditions are plotted as a function of periods in Figure 4. Data for the left and right hemispheres are presented separately in each panel. It is evident in the figure that CBFV was greater for the fast than for the slow event rate in both cerebral hemispheres and that in both hemispheres, CBFV declined over time in the two event rate conditions. These impressions were supported by a 2 (Event Rate) × 2 (Hemisphere) × 4 (Periods of Watch) mixed-ANOVA which revealed significant main effects for event rate, F (1, 18) = 4.82, $p < .05$, $\eta^2 = .21$, and periods of watch, F (2.12, 38.18) = 10.64, $p < .001$, $\eta^2 = .37$. The main effect of hemisphere and all interactions in the analysis were not significant, $p > .05$ in each case.

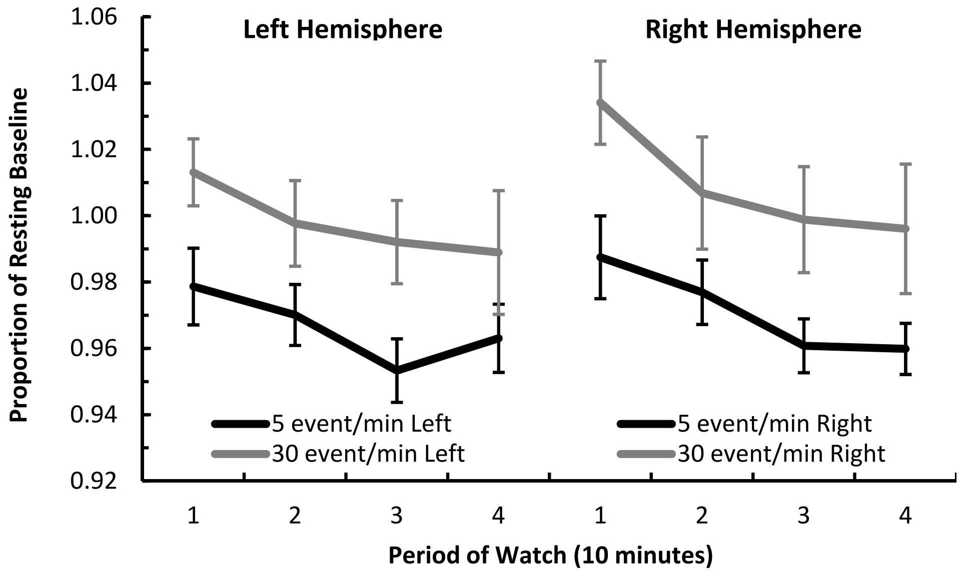


Figure 4. Hemovelocity scores for the two event rate conditions as a function of periods of watch. Data are plotted separately for each cerebral hemisphere. Error bars are standard errors.

Discussion

The present study tested several possibilities regarding the effects of background event rate on CBFV in a vigilance task. Specifically, it was anticipated that the overall level of CBFV would be greater in the context of a fast as compared to a slow event rate, that the decline in CBFV over time would be positively related to event rate, and that these effects would be lateralized to the right cerebral hemisphere. As expected, CBFV was greater with a fast event rate of 30 events/minute than with a slow rate of 5 events/minute. It is noteworthy that in addition to the rate of repetition of the background events in which critical signals occur in a vigilance task, the background events can also be varied in terms of whether they occur in a temporally regular or a temporally irregular manner. A recent study by Shaw et al. (2011) has found that CBFV is greater when the background events occur in an asynchronous or irregular than a synchronous or regular format. Taken together, the results of the present study and the experiment by Shaw and his associates demonstrate the importance of the background events in which critical signals are embedded on hemodynamics in vigilance performance.

Although the expectation regarding the relation between the overall level of CBFV and event rate was confirmed, the expectations regarding the temporal decline in CBFV and event rate and the lateralization of the CBFV effects were not supported by the data. Consistent with several earlier investigations (see Warm et al., 2008), CBFV declined significantly over time in this study, but the decline was independent of event rate and the CBFV effects associated with event rate and time on task were similar in both the right and left cerebral hemispheres. The absence of hemispheric effects is contrary to several earlier CBFV studies showing a right hemispheric system in the functional control of vigilance performance (Shaw et al., 2011; Warm et al., 2008). It is important to note, however, that the dominance of a right hemispheric control system in vigilance has been challenged by several recent studies showing bilateral activation in vigilance tasks (Helton et al., 2010; Schultz, Matthews, Warm, & Washburn, 2009; Shaw et al., 2011). Thus, while an overall right hemispheric system might be involved in the control of vigilance in some cases, in others, such as the present study, a cooperative interaction model (cf., Shaw et al., 2011) may best describe the role of cerebral functioning in the management of sustained attention. Future research is needed to discover the factors that determine brain symmetry and asymmetry in the maintenance of sustained attention under different experimental conditions.

A key feature of the relation between CBFV and vigilance is the parallel between CBFV and performance efficiency. That parallel was evident in the present study with regard to the temporal decrement in CBFV and the decline in the frequency of signals detection with time on watch. However, the CBFV/performance correspondence was not evident in regard to the effects of event rate. While CBFV was higher in the context of the fast as compared to the slow event rate, the two event rate conditions did not have differential effects upon the frequency of signal detection. At present, the reason for the lack of a performance difference between the two event rate conditions is not clear. However, to the extent that CBFV is an index of the utilization of information-processing resources, it would appear that in regard to event rate, it is a more sensitive index of task demand than performance efficiency. Even when performance differences were absent, the CBFV measure indicated that resource utilization was greater in the context of the fast as compared to the slow event rate.

References

Aaslid, R. (1986). Transcranial Doppler examination techniques. In R. Aaslid (Ed.), *Transcranial Doppler Sonography* (pp. 39-59). New York, NY: Springer-Verlag.

Coren, S., Ward, L. M., & Enns, J. T. (1999). *Sensation and perception* (5th ed.). Fort Worth, TX: Hartcourt-Brace.

Davies, D. R., & Parasuraman, R. (1982). *The psychology of vigilance*. London, UK: Academic Press.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Los Angeles, CA: Sage Publications Limited.

Helton, W. S., Warm, J. S., Tripp, L. D., Matthews, G., Parasuraman, R., & Hancock, P. A. (2010). Cerebral lateralization of vigilance: A function of task difficulty. *Neuropsychologia*, *48*, 1683-1688.

Hitchcock, E. M., Warm, J. S., Matthews, G., Dember, W. N., Shear, P. K., Tripp, L. K., Mayleben, D.W., & Parasuraman, R. (2003). Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task. *Theoretical Issues in Ergonomics Science, 4*, 89-112.

Navon, D. (1984). Resources: A theoretical soupstone? *Psychological Review, 91*, 216-234.

Oldfield, R. C. (1971). The assessment and analysis of handedness. *Neuropsychologica, 9*, 97-113.

Parasuraman, R., Warm, J. S., & Dember, W. N. (1987). Vigilance: Taxonomy and utility. In L. S. Mark, J. S. Warm, & R. L. Huston (Eds.), *Ergonomics and human factors: Recent research* (pp. 11-32). New York, NY: Springer-Verlag.

Schultz, N. B., Matthews, G., Warm, J. S., & Washburn, D. A. (2009). A transcranial Doppler sonography study of shoot/don't-shoot responding. *Behavior Research Methods*, *41*, 593-597.

Shaw, T., Finomore, V., Warm, J., & Matthews, G. (2011). Effects of regular or irregular event schedules on cerebral hemovelocity during a sustained attention task. *Journal of Clinical and Experimental Neuropsychology*, *34*, 57-66.

Shaw, T. H., Warm, J. S., Finomore, V. S., Tripp, L., Matthews, G., Weiler, E., & Parasuraman, R. (2009). Effects of sensory modality on cerebral blood flow velocity during vigilance. *Neuroscience Letters*, *461*, 207-211.

Toole, J. F. (1984). *Cerebrovascular disorders* (3rd ed.). New York, NY: Raven Press.

Tripp, L., & Warm, J. S. (2007). Transcranial Doppler sonography. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The brain at work* (pp. 82-94). New York, NY: Oxford University Press.

Warm, J. S., Finomore, V., Vidulich, M. A., Funke, M. (in press). Vigilance: A perceptual challenge. In R. R. Hoffman, P. A. Hancock, R. Parasuraman, J. L. Szalma, & M. Scerbo (Eds.), *The handbook of applied perception research*. New York, NY: Cambridge University Press.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors, 50*, 433-441.

THE USE OF FUNCTIONAL NEAR INFRARED SPECTROSCOPY (FNIRS) TO ASSESS COGNITIVE WORKLOAD OF AIR TRAFFIC CONTROLLERS

Hyun Woo[1], Ben Willems[1], Joshua Harrison[2], Kurtulus Izzetoglu[2], Sehchang Hah[1], & Hasan Ayaz[2]
[1]Federal Aviation Administration
Atlantic City International Airport, NJ
[2]Drexel University
Philadelphia, PA

The performance of a Certified Professional Controller (CPC) can have a critical impact on safety. A specific concern is that a high cognitive load has been associated with performance decrement. Thus, it is important to continuously monitor and accurately assess CPC cognitive load. The purpose of the study was to evaluate the effectiveness of Conflict Resolution Advisory (CRA), automation which provides CPCs with resolutions to avoid conflicts. In this study, we used functional Near Infrared Spectroscopy (fNIRS) to index cognitive workload of 12 CPCs from En Route centers. Results indicate that fNIRS measures were sensitive to air traffic level, but we did not find significant differences across CRA conditions. In addition, we conducted analysis on fNIRS data time-locked to selected events such as clearance commands. With this event-related analysis, we found differences among CRA conditions for different events. The findings indicate that fNIRS can be a potential objective workload measurement tool in the air traffic control domain.

The Federal Aviation Administration (2012) predicts that the amount of air travel will double within two decades. Increased air traffic will lead to increased mental workload for Certified Professional Controllers (CPC). Since performance decrement has been associated with high mental workload (Strayer & Drews, 2006), safety becomes a concern when CPCs' mental workload increases.

It is essential to accurately measure mental workload in order to effectively address safety issues. Mental workload can be indexed using many different methods including secondary task performance measures, subjective measures, and physiological measures. Secondary task measurement and subjective measures can be obtrusive and biased. Physiological measures, however, allow an objective and continuous indexing of mental workload. Neurophysiological measures such as electroencephalogram (EEG) or functional magnetic resonance imaging (fMRI) provide more direct measurements of brain function. Similar to fMRI, functional Near Infrared Spectroscopy (fNIRS) is a hemodynamic measure. fNIRS uses optical signaling to provide biomarkers, such as oxygenated and deoxygenated hemoglobin concentration changes, to index cognitive workload. Unlike fMRI and EEG, fNIRS users' movements are not strictly confined, allowing the device to be a more practical candidate for real-world applications.

In this study, we used fNIRS in an En Route simulation to examine the effectiveness of a decision aid, Conflict Resolution Advisory (CRA). CRA provides controllers with possible resolutions to avoid conflicts in order to reduce their workload during a decision making stage. First, we compared cognitive workload of different air traffic levels. We then compared controller positions and CRA conditions. Finally, we conducted event-related analyses to examine epochs only around decision making timeframes.

Methods

Participants

Twelve En Route CPCs (2 females, mean age 40.7 ± 10.1 years) participated in the study. All participants were active controllers with a mean experience of 14.6 ± 11.3 years and normal or corrected-to-normal vision.

Data Acquisition

We used Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE) to simulate the En Route Automation Modernization (ERAM) system. DESIREE was time-synchronized with other data acquisition tools such as fNIRS and a subjective workload assessment tool. DESIREE simulated two active high altitude sectors, 20 and 22, of Kansas Center (ZKC). In the scenarios, the air traffic steadily increased from 33% to 150% of the Monitor Alert Parameter (MAP) value. DESIREE recorded all user data, including mouse clicks, keyboard inputs, and controller commands.

We used Workload Assessment Keypads (WAK) to measure subjective workload during the experiment. Participants were prompted to input an index of their workload on 10-point scale every two minutes, where “1” reflected low workload and “10” reflected high workload. Participants had maximum of 20 seconds to respond, and after 20 seconds, responses were marked as “missed.”

We used two continuous wave fNIRS systems developed at Drexel University (Philadelphia, PA), manufactured and supplied by fNIRS Devices LLC (Potomac, MD; www.fNIRSdevices.com) to index prefrontal cortex activity. There were 16 sensors approximately separated by 2.5 cm with approximately 1.25 cm penetration depth. The device collected data at a rate of 2 Hz with Cognitive Optical Brain Imaging (COBI) Studio software (Drexel University) for data collection.

Procedure

The study lasted three weeks with four controllers per week. The first day was devoted to training, the following three days were devoted to experimental sessions, and the last day was devoted to travel. Each experimental session had four participants where two participants were assigned to Sector 20 while the other two were responsible for Sector 22. Within each sector, each participant was assigned to either radar console (R-side) or radar associate position (D-side). Each of the two participants on Sector 20 wore an fNIRS below an eye-tracking device while participants on Sector 22 wore an electroencephalogram (EEG) device. Participants rotated positions according to a predefined schedule. In this paper, we will limit our analysis and discussion to fNIRS data and participants on Sector 20.

We tested three conditions to examine the effects of CRA automation. “No CRA” condition did not have CRA on both R-side and D-side, “CRA on D” condition had CRA implemented only on D-side, and “CRA on B” condition had CRA implemented on both R-side and D-side. There were three runs of same condition for each day of experimental sessions. The order of CRA conditions were counterbalanced each week.

Preprocessing of fNIRS data

We used Matlab (The MathWorks Inc., MA, USA) software to analyze the fNIRS data. We used a sliding motion artifact removal algorithm (SMAR) and linear phase filter with cutoff frequency of 0.14 Hz to filter the data. This helped to remove any detectable motion artifacts attenuate high frequency noise, and compensate for respiration and cardiac cycle effects (Ayaz, Izzetoglu, Shewokis, & Onaral, 2010;

Ayaz et al., 2012). We excluded any saturated channels and channels with less than 75% of valid data points. Finally, we calculated the blood oxygenation change relative to the first two minutes baseline using the modified Beer-Lambert Law.

Results and Discussion

In our analysis, we used the Jeffrey-Zellner-Siow (JZS) Bays factor *t*-tests via a Web-based program (at pcl.missouri.edu) developed by Rouder, Speckman, Sun, & Morey (2009). Compared to the traditional statistical method of the Null Hypothesis Significance Testing (NHST), Bayesian model comparison allows researchers to state evidence for the null hypothesis. For example, in NHST, researchers can reject the null hypothesis, or fail to reject the null hypothesis, but are not allowed to state the evidence for the null hypothesis. However, the Bayesian model compares probability of the null over the alternative where Bayes Factor (BF) above 1 indicates evidence for the null while BF below 1 indicates evidence for the alternative. It gives the anecdotal, substantial, or strong evidence that two conditions are not different. Benefits of Bayes analysis are described in detail in Dienes (2011).

Air Traffic Level

Previous simulation studies have found relationship between traffic volume and oxygenation levels measured by fNIRS (Ayaz et al., 2011). Similarly, we tested the relationship between oxygenation levels and traffic volume. We defined low, medium, and high traffic by 7 to 13, 14 to 20, and 21 to 27 aircraft, respectively. We used the maximum number of aircraft under control by each participant and the oxygenation change of 10 seconds before and 10 seconds after the aircraft count. There was a significant effect for air traffic levels between low and medium traffic, $t(4)=5.61$, $p<.005$, $BF=0.07$, medium and high traffic, $t(4)=5.54$, $p=.005$, $BF=0.07$, and low and high traffic, $t(4)=8.76$, $p<.001$, $BF=.02$. As shown in Figure 1, results still showed significant results when controller position and CRA condition was taken into account.

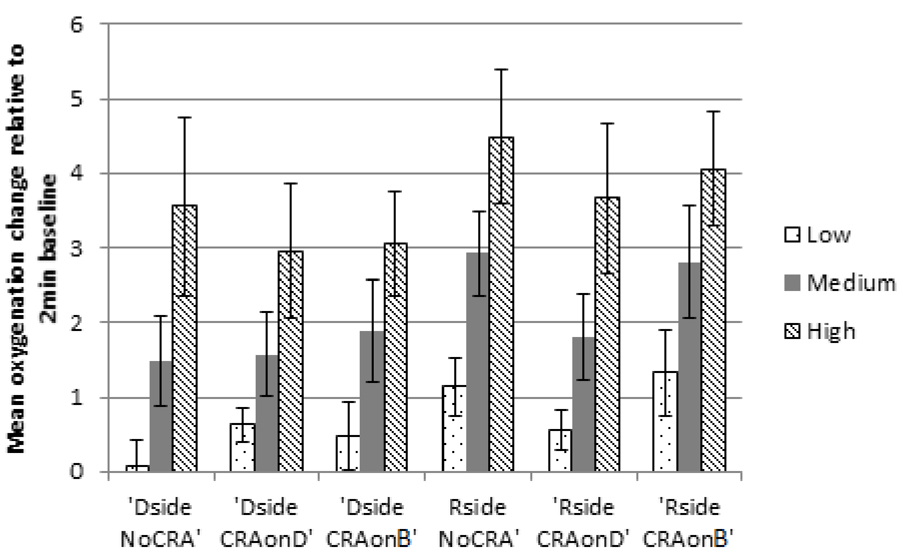


Figure 1. Mean oxygenation change relative to two-minute baseline for three levels of aircraft volume. Error bars represent SEM. With exception of 'D-side, CRA on D,' all other conditions show significant oxygenation difference between low vs. medium traffic and low vs. high traffic. Only 'R-side No CRA' and 'R-side CRA on B' show significant oxygenation differences between medium vs. high traffic.

Controller Position Comparisons

Paired *t*-test and Bayes factor analysis showed anecdotal evidence for no difference between mean oxygenation change for controller positions D-side and R-side for No CRA ($t(5) = 1.01$, p = .36; *BF* = 2.23), CRA on D ($t(7) = 1.55$, p = .17; *BF* = 1.47), and CRA on B ($t(4) = .46$, p = .67; *BF* = 2.93).

CRA Condition Comparisons

Paired *t*-test and Bayes factor analysis showed no support for differences in mean oxygenation change between conditions for both D-side and R-side. Bayes factor (BF) shows substantial evidence for the null ($3.0 < BF < 10.0$) providing evidence that there is no difference between CRA conditions as shown in Table 1.

Table 1.
Summary t-test and Bayesian statistics for CRA condition comparison.

	D-side			R-side		
CRA condition	*No-D*	*No-Both*	*D-Both*	*No-D*	*No-Both*	*D-Both*
p-value	0.68	0.72	0.84	0.67	0.82	0.47
Sample size	7	7	7	7	6	7
t-value	0.44	0.38	0.21	0.45	0.24	0.77
Bayes Factor (*BF*)	3.38	3.46	3.63	3.37	3.39	2.84

We expected to find differences among CRA conditions when we compared CRA conditions for each blocks of aircraft traffic level (low, medium, high). However, we did not find significant difference among condition, and Bayes analysis shows anecdotal evidence for the null ($BF > 1$).

Event-Related Analysis

We were interested in how controllers' workload changed before and after they executed a clearance (i.e., command). We hypothesized that oxygenation level would rise before a clearance issuance because controllers would need to exert mental resources in order to make a clearance. For instance, controllers would evaluate the situation, consider different options, make a decision, and execute a clearance. We chose to look at 10 seconds (s) before and after a clearance because previous studies have shown that hemodynamic response evolves over a 10s to 12s period or less (Miezin, Maccotta, Ollinger, Petersen, & Buckner, 2000; Bunce, Izzetoglu, Izzetoglu, Onaral, & Pourrezaei, 2006; Kim, Richter, & Ugurbil, 1997). Ten seconds after each clearance would have lower oxygenation level than 10s before a clearance because a decision was already made. We examined three common commands that controllers use to control traffic: change in altitude, change in heading, and change in speed. However, we did not see any difference between the times before and after the clearance issuances (see Figure 2). This may be because of the complexity of air traffic control. Within 10s, controllers are executing many commands and performing multiple tasks. We only examined 10s before a command for further analysis because of this lack of difference.

Even though we did not see any overall difference in oxygenation levels across different CRA conditions, we expected to see a difference when we examined the epochs around a command. For example, if CRA effectively assists the controllers in making decisions and lowering their workload, the oxygenation level of the conditions where CRA was available would be lower than the No CRA condition. By examining the epochs around a command, all other noise would be averaged out. As shown in Table 2, we found some oxygenation difference among CRA conditions for different commands. For example, there is a decisive evidence for difference between CRA on D and CRA on Both on D-side controllers when giving altitude change command, and between No CRA and CRA on Both on R-side controllers when giving speed change command.

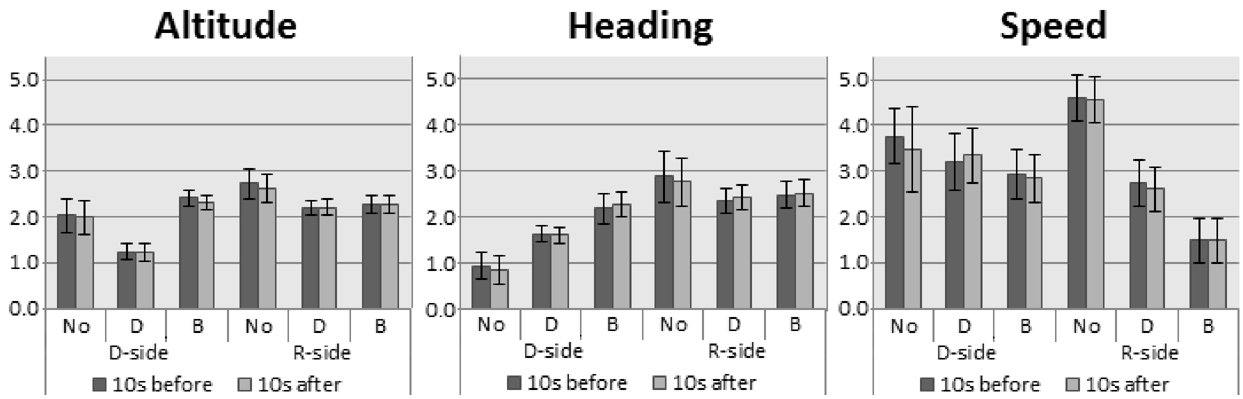


Figure 2. Oxygenation change 10s before and 10s after clearances issued by using the datablock. Error bars represent SEM.

Table 2.
Summary t-test and Bayesian statistics for DB oxygenation 10s before commands

Command	Position	Condition	p-value	df	*t*-value	BF
Altitude	D-side	No-D	0.055	81	1.95	1.41
		No-Both	0.329	81	0.98	5.7+
		D-Both	0.00	106	5.12	0****
	R-side	No-D	0.154	50	1.45	2.66
		No-Both	0.227	50	1.22	3.91+
		D-Both	0.786	78	0.27	8.91+
Heading	D-side	No-D	0.059	23	1.99	0.91
		No-Both	0.007	23	2.98	0.11*
		D-Both	0.141	48	1.5	2.33
	R-side	No-D	0.388	17	0.89	3.35+
		No-Both	0.52	17	0.66	4.06+
		D-Both	0.732	42	0.35	6.08+
Speed	D-side	No-D	0.536	3	0.7	2.35
		No-Both	0.332	3	1.15	1.71
		D-Both	0.724	19	0.36	4.31+
	R-side	No-D	0.015	15	2.75	0.2*
		No-Both	0.00	15	5.22	0****
		D-Both	0.083	15	1.85	0.98

Note. *substantial evidence for difference, **strong evidence for difference, ***very strong evidence for difference, ****decisive evidence for difference, +substantial evidence for *no* difference

Conclusion

This paper demonstrated use of fNIRS to index mental workload in a high fidelity simulation of En Route air traffic environment. Although we did not find any statistical significance of mental workload among CRA conditions, we were able to find differences in air traffic level and event-related analysis. A limitation of this study is that we were not able to conduct an analysis for each of the 16 channels due to small sample size and noisy data. In addition, it should be noted that in our study, the fNIRS indexes mental workload only from the prefrontal cortex. It is a possibility that with high workload, resources may be reallocated to different part of the brain which fNIRS cannot measure. Future analysis should include correlates of fNIRS measures with other measures of mental workload such as subjective measurements and performance measurements. In addition, further research on event related analysis with fNIRS can be an interesting method to examine the pattern of oxygenation related with mental workload. Even though not reported here, future fNIRS studies should conduct an analysis by each channel to examine spatial allocation of oxygenation in prefrontal cortex.

Acknowledgements

The authors would like to thank the Human Factors Branch (ANG-E25), the Concept & Systems Integration Branch (ANG-E14), and their contractors. We also would like to thank Sonia Alvidrez and Rob Bastholm for data collection and comments on the manuscript, the Concept Analysis Branch (ANG-C41) for collecting the fNIRS data, and Drexel University for their help on fNIRS data.

References

Ayaz H., Izzetoglu, M., Shewokis, P., & Onaral, B. (2010). Sliding-window motion artifact rejection for functional near-infrared spectroscopy. Proceedings from IEEE EMBS: *32nd Annual International Conference of the IEEE EMBS.* Buenos Aires, Argentina.

Ayaz, H., Willems, B., BUnce, S., Shewokis, P., Izzetoglu, K., Hah, S., …Onaral, B. (2011). Estimation of cognitive workload during simulated air traffic control using optical brain imaging sensors. In Schmorrow, D. & Fidopiastis, C. (Eds.), *Foundations of Augmented Cognition: Directing the Future of Adaptive Systems* (pp. 549-558). Springer Berlin/Heidelberg.

Ayaz, H., Cakir, M., Izzetoglu, K., Curtin, A., Shewokis, P., Bunce, S., & Onaral, B. (2012). Monitoring expertise development during simulated UAV piloting tasks using optical brain imaging. Proceedings from *Aerospace Conference, 2012 IEEE.* Big Sky, MT.

Bunce, S., Izzetoglu, M., Izzetoglu, K., Onaral, B., & Pourrezaei, K. (2006). Functional near-infrared spectroscopy: An emerging neuroimaging modality. *IEEE Engineering in Medicine and Biology Magazine 25*(4). 54-62.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*(3), 274-290.

Federal Aviation Administration. (2012). *FAA aerospace forecast: Fiscal years 2012-2032.* Retrieved from http://www.faa.gov/data_research/aviation/.

Kim, S., Richter, W., & Ugurbil, K. (1997). Limitations of temporal resolution in functional MRI. *Magnetic Resonance in Medicine, 17.* 631-636.

Miezin, F., Maccotta, L., Ollinger, J., Petersen, S., & Buckner, R. (2000). Characterizing the hemodynamic response: Effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *NeuroImage, 11.* 735-759.

Rouder, J., Speckman, P., Sun, D., & Morey, R. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16* (2). 225-237.

Strayer, D., & Drews, F. (2006). Cell-phone-induced driver distraction. *Current Directions in Psychological Science, 16*(3). 128-131.

An Integrated Neuroergonomic Assessment of In-Flight Pilot Workload

Steven D. Harbour[1], James C. Christensen[1, 2], Justin R. Estepp[2], and Tyron M .Gray[2]
Northcentral University[1] and Air Force Life Cycle Management Center[1], and the Air Force Research Laboratory[2]
Prescott, Arizona[1] and Wright Patterson Air Force Base, Dayton, Ohio[1, 2]

This study accomplished an exploration of a workload protocol and a preliminary assessment of pilot workload and situation awareness in-flight during various phases of airborne operation on a tactical airlift aircraft. Initial comparisons were made between the Head Down Display (HDD) which was previously certified by the FAA as a Primary Flight Display (PFD), and the original Head Up Display (HUD) configuration which has not been endorsed as a PFD. Quantitative results were produced to aid in resource decision making regarding cockpit systems for this aircraft. Low and no cost improvements were identified and recommendations were implemented. This work refined a neuroergonomic protocol for workload assessment.

In flight operations, increasing workload at some point decreases pilot or operator attention, leading the way to mishaps such as controlled flight into terrain (CFIT) (Wickens & McCarley, 2008). Consequently, workload assessment is a critical and ongoing issue despite many years of research (e.g., Tsang & Vidulich, 2006). Measuring aircraft pilot or operator workload in an aerospace context presents unique challenges. For example, subjective self-report measures have traditionally been used to assess workload and SA in the cockpit (AFFSA, 2007) post-flight. This time delay, often more than one hour, can degrade recollection accuracy (Baddeley, 2006; Moroney et al., 2005; Southwick, Morgan, & Hazelett, 2007). While simulations may be paused for collection, actual flight operations preclude this. Therefore, subjective workload data from actual flight may be error prone, causing the workload assessment to be problematic. In addition, the authors have observed that military pilots typically may not be completely forthcoming in self-reporting workload. Therefore, non-intrusive, objective measures have long been sought. The present work sought to test and verify the utility of combined subjective and heart-rate based assessment in tactical airlift flights.

METHODOLOGY

The methods were chosen to be minimally intrusive to both the pilot and the mission. The instruments used have been well-established in previous research; this research was not intended to develop new measures but instead to establish a workload assessment protocol and collect real-world data. Subjective (survey) instruments included the Bedford Workload Scale and China Lake Situation Awareness Scale. These were administered to the pilots post flight. In addition heart rate was measured with the use of three-lead ECG (Hankins & Wilson, 1998; Nickel & Nachreiner, 2003.

The objective measures obtained from ECG included heart rate (HR) and heart rate variability (HRV). Both measures have been extensively studied, and in this study are taken to be broadly reflective of stress experienced by the pilot. To maximize the ecological validity of this study, all data collection was accomplished with actual flight operations. To avoid disrupting training activities, experimenters could not direct the flight profile for each sortie; relevant segments were extracted post hoc. Consequently this study is a quasi-experimental design.

Participants and Procedures

Participants. Eleven participants were recruited from the Ohio Air National Guard population and ranged in ages from 25 to 46. All of the participants had flying experience and were qualified in the aircraft. Participants flew using either the HDD or both the HDD and HUD (See figures 1, 2, 3, 4, & 5). The experimental protocol was reviewed and approved by the Air Force Research Laboratory's IRB and the Aeronautical Systems Center's Flight Safety Review Board. As the study was part of their normal duties, no additional compensation was provided.

Tasks. Aircrew piloted a tactical airlift aircraft flying normal scheduled training missions that consisted of some combination of takeoff, cruise, low level, airdrop, and assault landing phases.

Equipment. The military aircraft had an L-3 Communications HDD and a Rockwell Collins Flight Guidance Systems HUD. The electrocardiographic data was collected using a Vitaport system (Temec Instruments B.V., Kerkrade, Netherlands), which is a small, portable pilot-worn physiological data collection system with onboard digital data storage. The three leads were placed on left and right clavicles and sternum; impedances were verified at or below 40 kOhm.

Procedures. After completing written informed consent, pilots were instrumented for ECG. An experimenter accompanied the flights, and used dedicated marker channels along with written notes to time stamp and categorize flight segments and events of interest. A typical sortie was approximately three hours in duration, with changes in pilot flying and maneuvers being practiced. Post-flight, pilots were asked to complete a set of subjective measures for each segment.

Figure 1. Head Down Displays (HDDs).

Figure 2. Pilot's (HDDs).

Figure 3. Co-Pilot's HDDs.

Figure 4. Pilot's HUD.

Figure 5. Co-Pilot's HUD.

RESULTS

Due to the quasi-experimental nature of this study, there are different numbers of pilots represented in each comparison; this will be noted throughout. Due to the unequal numbers and multiple comparisons, omnibus statistics have not been generated. Multiple sorties or repeated maneuvers were averaged together for each pilot and maneuver type; the plotted means and standard errors reflect a single value from each pilot. All data presented here are drawn from the pilot flying. The HR and HRV data were analyzed by extracting 5-minute segments during each maneuver. ECG data were bandpass filtered from .4 Hz to 30 Hz. R wave peaks were marked using QRStool based on threshold detection followed up with visual inspection and correction. Interbeat intervals were exported for subsequent analyses of average HR and SDNN (in this case, standard deviation of the R-R interval).

The first comparison involved 3 pilots, each of whom completed 2 sorties utilizing both the HDD and original HUD configuration to conduct multiple non-precision instrument approaches, uncoupled from the autopilot (Fig. 1).

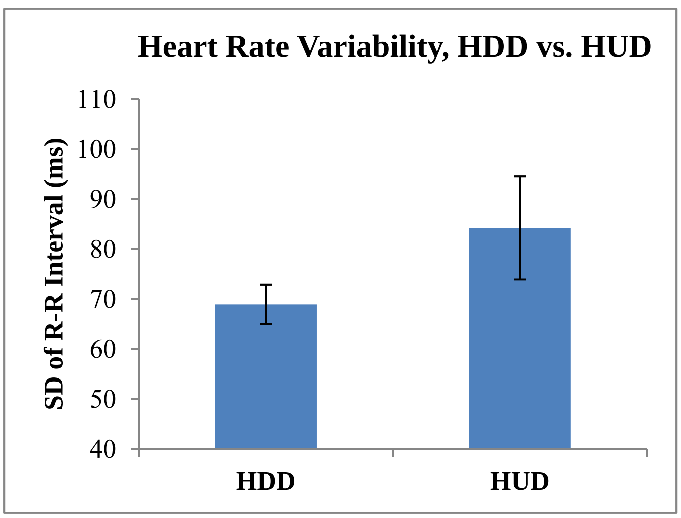


Figure 1. Average SDNN for instrument approaches conducted solely heads down (HDD) and with the head-up display available (HUD). Bar height represents the mean, and error bars are one standard error of the mean. Note that higher values are associated with decreased stress and workload.

Based on Figure 1, it appears that the use of the HUD lowers workload. This was also reflected to some degree in the self-report surveys of workload and SA (Figs. 2 and 3). However, note that the self report measures exhibit compression and floor effects – self rated workload is very low and SA very high in both cases.

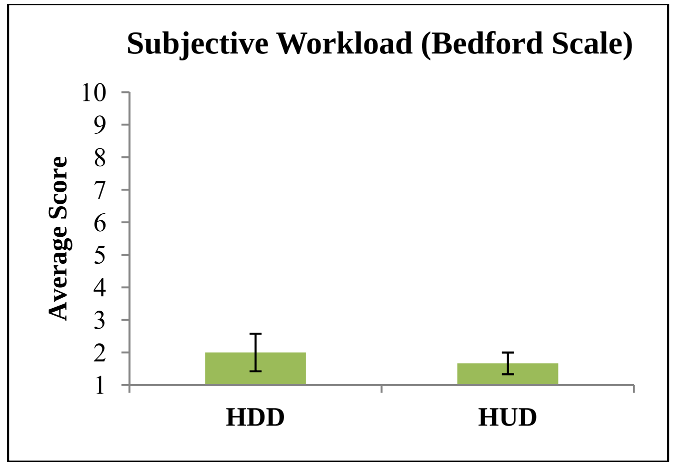


Figure 2. Average subjective workload ratings, on the Bedford 1 to 10 scale. Higher values correspond to higher workload.

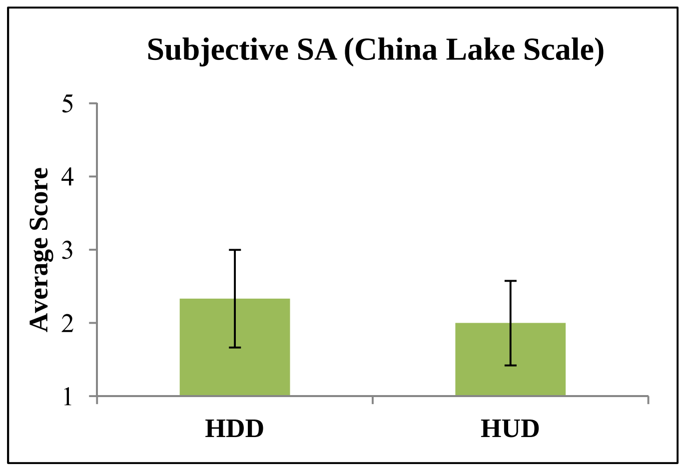


Figure 3. Average subjective situation awareness (SA) ratings, on the China Lake 1 to 5 scale. Higher values correspond to lower SA.

A second comparison was drawn from low-level air drop maneuvers. 5 minute segments were generated by counting back from the release point; the comparison for reference was drawn from data during circling back to begin another air drop. A new set of 3 pilots conducted air drops, with each contributing one average value based on 3 to 4 repetitions of the maneuver (Figs. 4-6).

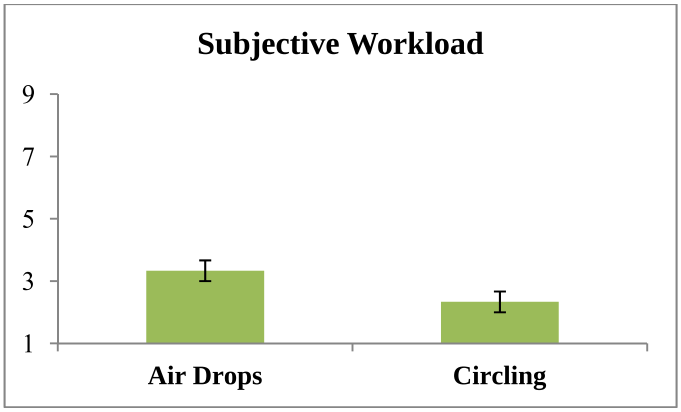


Figure 4. Subjective workload (Bedford scale) during air drops and circling. Higher values correspond to higher workload. Note: circling is at a one.

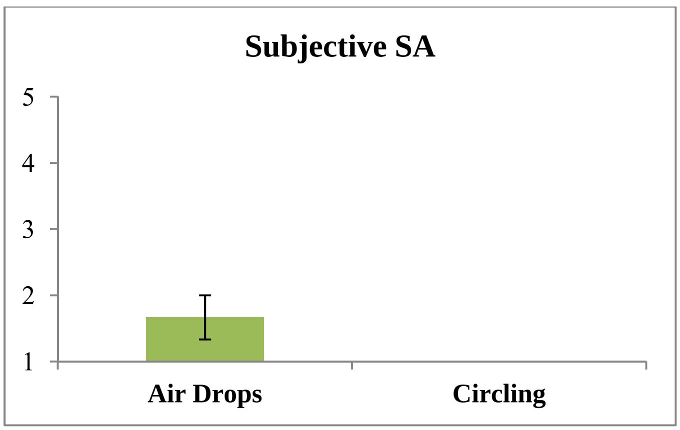


Figure 5. Subjective situation awareness (China Lake scale) during air drops and circling. Lower values correspond to better SA.

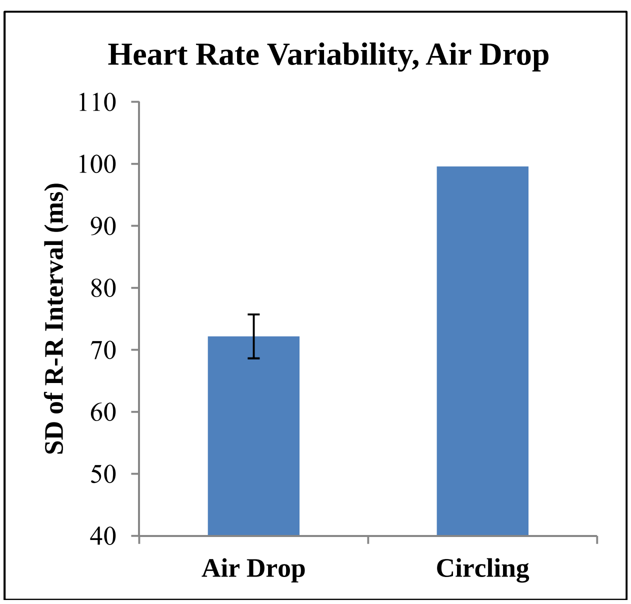


Figure 6. Average SDNN for low-level air drops as compared with circling back to conduct another drop. Variance was negligible during circling, hence the missing error bar. Higher values correspond to decreased stress/workload.

Assault landings (steep, high-speed approach) resulted in the lowest HRV measures (Fig. 7) of all the maneuvers tested. Eight pilots completed at least five assault landings each. The subjective measures again exhibited compression with a mean workload of 2.7 out of 10, and SA 2 out of 5. Figure 7 was shared with the instructor pilots from the participating units; they confirmed that the relative ordering of maneuver types matches their own opinions.

Given the concerns in the literature regarding HRV measures of workload, future analysis will continue with HR measures as well. One example of HR reactivity to a significant event (air traffic announced by ATC in proximity to the aircraft) is presented in Figure 8.

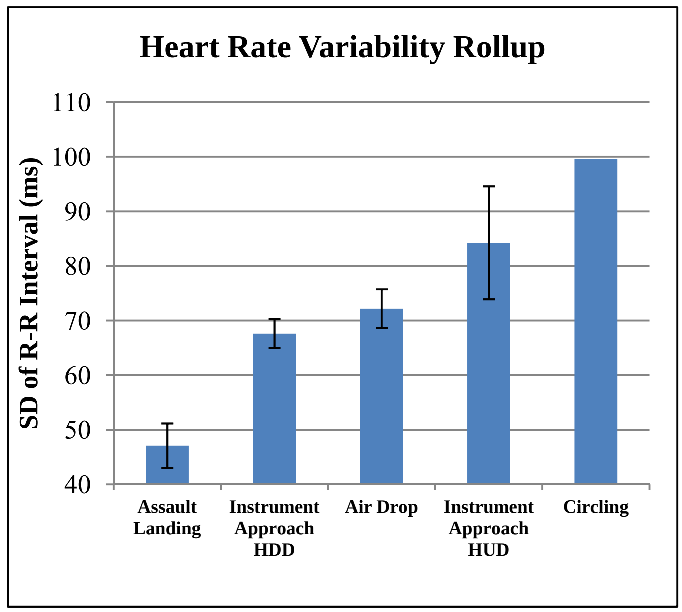


Figure 7. Average SDNN for all the maneuvers analyzed to date. Higher values correspond to decreased stress/workload.

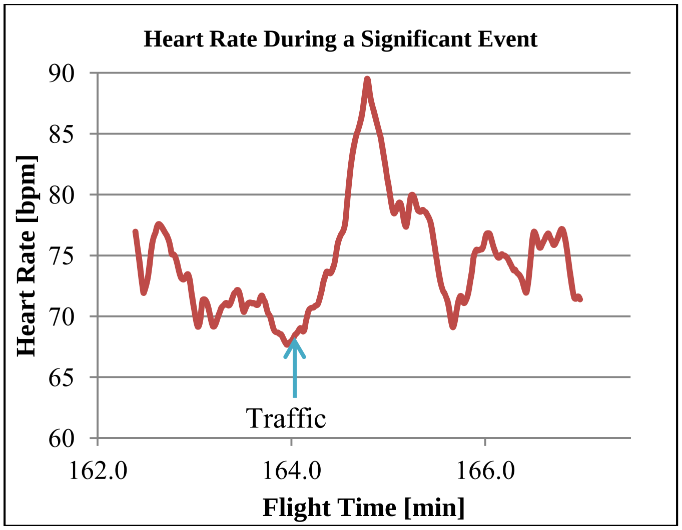


Figure 8. Heart rate (HR, in beats per minute) over a 5 minute flight segment taken centered on a significant event, specifically air traffic that required maneuvering to avoid. HR is based on IBIs, smoothed with a 30-sample moving window average (29 sample overlap).

DISCUSSION

The primary findings of this study address firstly the workload experienced in tactical airlift flight operations, and secondly methodological concerns regarding single measure workload assessment with military aircrews.

Similar to Wilson and Fisher (1991), we observed noticeable differences in heart activity between different flight segments. Based on both increases in heart-based measures of workload and subjective ratings, assault landings, air drops, and instrument approaches flown without a HUD were identified as areas of concern with regard to pilot stress and workload. The following recommendations were made.

Recommendation I: Workload during Airdrops was found to be high. Autopilot usage would significantly reduce workload. As a result this recommendation was implemented.

Recommendation II: For assault landings, instrument approaches, and low-level flight a HUD, particularly a certified PFD HUD could reduce workload.

The authors who are rated pilots noted that their own observations of pilot workload in flight were not consistent with the subjective ratings. For example, pilots rated hand-flying a low level air drop as 3 out of 10 on the Bedford scale; the anchor text is "Enough spare capacity for all desirable additional tasks". We observed channelized attention focused on minimizing lateral deviations, with consequent mistaken radio calls and lack of attention to secondary displays. Our observations were more consistent with 6 or perhaps 7 on this scale. While there may be perceived pressure to underrate workload despite assurances of no negative repercussions, we also feel that pilots may misperceive their workload in the absence of a highly visible error. If this is indeed a systematic feature of subjective ratings taken from this type of population and task, it further underlines the importance of integrating multiple measurement types in conducting workload studies.

Impact:

The results of this experiment show that a physiological workload assessment reveals decreased workload when using even the original HUD during instrument approaches. There are ongoing concerns about workload during certain phases of flight; this study is providing input to USAF cockpit system upgrade decisions.

CONCLUSIONS

The results presented here indicate that an integrated neuroergonomic protocol and method for workload assessments of pilots and operators of aerospace vehicles in-flight is appropriate. Adding the objective physiological measurement of heart rate augments the veracity and robustness of the findings. With these considerations in mind, workload assessments using both subjective and objective measures, interdependently, provide a more complete picture of pilot workload and SA than does the use of either alone. More research in this area needs to be accomplished, such as adding Electroencephalogram (EEG) data collection, and testing for the effects of varying spatial locations of information displays (ID). Flying and utilizing a display primarily utilizes vision. Therefore, in order to answer the theoretical questions of why or how workload is changed due to ID design, further research should be accomplished investigating individual differences in visual attention and perception as mediating neurocognitive characteristics, thus contributing to further refinements in theories of workload and attention.

REFERENCES

Adams, S.R., Kane, R. & Bates R. (1998). Validation of the China Lake Situational Awareness scale with 3D SART and S-CAT. China Lake, CA: Naval Air Warfare Center Weapons Division (452330D).

Air Force Flight Standards Agency. (2007). *White paper: Primary flight reference endorsement process.* Washington, DC: Author.

Baddeley, A. D. (2006). Working memory: An overview. In S. J. Pickering (Ed.), *Working memory and education* (pp. 3–31). Burlington, MA: Elsevier.

Hankins, T.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, 360-367.

Moroney, W.F., Biers, D.W., Eggemeier, F.T., and Mitchell, J.A. (1992). A comparison of two scoring procedures with the NASA task load index in a simulated flight task. In *Proceedings of the 1992 National Aerospace and Electronics Conference* (pp. 734-740). New York: Institute of Electrical and Electronics Engineers.

Nickel, P., & Nachreiner, F. (2003). Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload. *Human Factors*, 45(4), 575-590.

Roscoe, A.H., & Ellis, G.A. (1990). A subjective rating scale for assessing pilot workload in flight: A decade of practical use. Bedford, UK: Royal Aerospace Establishment.

Southwick, S., Morgan, C., and Hazelett, G. (2007). Memory for trauma: Accurate, inaccurate, or just plain false? Retrieved May 26, 2010, from Yale University Educational Videos On Line. Web site: http://www.med.yale.edu/psych/education/videos.html

Tsang, P.S., and Vidulich, M.A. (2006). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (3rd ed., Chap. 9). New York: Wiley.

Wickens, C. D., & McCarley, J. S. (2008). *Applied attention theory*. Boca Raton, FL: Taylor and Francis Group.

Wilson, G.F., & Fisher, F. (1991). The use of cardiac and eye blink measures to determine flight segment in F4 crews. *Aviation, Space, and Environmental Medicine, 62,* 959-961.

Perceptual and Adaptation Implications with Display 3-D Spatial Location: Retrofit of HUD on a Tactical Airlift Platform

Steven D. Harbour[1], Jeffery A. Hudson[2], and Gregory F. Zehner[3]
Northcentral University & Air Force Life Cycle Management Center[1], Infoscitex[2] , and 711 HPW/HPO[3]
Prescott, Arizona[1] and Wright Patterson Air Force Base, Dayton, Ohio[1, 2, 3]

The retrofitting of a cockpit with a Head-Up-Display (HUD) raises potential accommodation and perceptual issues for pilots that must be addressed. For maximum optical efficiency, the goal is to be able to place every pilot's eye into the HUD Eye Motion Box (EMB) given a seat adjustment range. Initially, the Eye Reference Point (ERP) of the EMB should theoretically be located on the aircraft's original cockpit Design Eye Point (DEP) while horizontal and vertical seat adjustment would allow pilots to position their eyes inside the EMB. However, human postures vary, and HUD systems may not be optimally placed. In reality there is a distribution of pilot eyes around the DEP (which is dominant eye dependent) therefore this must be accounted for in order to obtain appropriate visibility of all of the symbology based on photonic characteristics of the HUD. Pilot size and postural variation need to be taken into consideration when positioning the HUD system to ensure proper vision of all HUD symbology in addition to meeting the basic physical accommodation requirements of the cockpit. The innovative process and data collection methods for maximizing accommodation and pilot perception on a new "tactical airlift" platform are discussed as well as the related neurocognitive factors and the effects of information display design on cognitive phenomena.

In commercial aviation 983 accidents occurred between 1980 and 2007, involving multiengine jet aircraft that weighed 12,500 pounds or more with cockpits that did not contain HUDs (Flight Safety Foundation, 2009). From information obtained in flight simulators, it is believed that up to 73% of those accidents could have been prevented had a HUD been installed in the cockpit (Flight Safety Foundation, 2009; Kim, 2009). That study made those estimations with the assumption that the HUD was optimally placed in the cockpit. It also assumed that the HUD EMB is matching the Cockpit DEP, and is within the pilot Line-of-Sight (LOS), which often times is not the case once the HUD system is actually installed in the aircraft, especially if it did not originally come with a HUD (Hudson, Zehner, Harbour, & Whitehead, 2011). A method should be established to ensure that the HUD is ideally spatially located in the cockpit to reduce pilot workload and increase situation awareness (Harbour, Christensen, Estepp, & Gray, in press).

PURPOSE

The purpose of our study was to conduct an initial anthropometric and ergonomic cockpit assessment in order to measure the pilot interface with cockpit functions. This study specifically mapped real Pilot and subject Eye Box locations in the cockpit and the HUD Pilot Eye Motion Box for the prototype HUD installation. Both were digitized using a FARO arm, which was also used to reverse engineer the entire cockpit and put into Computer Aided Design (CAD) software. These geometric data are necessary to consider the perceptual effects onto mental Workload (WL) and Situation Awareness (SA).

METHODOLOGY

This study served as a "first look" to assess the major accommodation problems associated with a HUD installation. A complete accommodation evaluation that determines a percentage accommodation for the flying population, and which also quantifies reach and clearance minimums and maximums, will be done at a later date. For this study, five women and fifteen men were selected based on their Sitting Eye Heights, which ranged from 29.2" to 35.0," covering the entire male range and most of the female range reported in the Aircrew Sizing Survey (ACSS, Table 1 below).

Table 1. USAF Aircrew Sizing Survey: 2009-2011			
SITTING EYE HEIGHT	Mean	5th %ile	95th %ile
Males	32.4"	30.5"	34.4"
Females	30.4"	28.3"	32.2"

Their Mid-Pupil location was mapped in 3D Space with a FARO arm after he or she visually "lined up" on original cockpit "design eye spheres" (hence, they were positioned at their perceived DEP). In addition, seat position and eye location for the four in-aircraft qualified pilot subjects were recorded where he or she would actually fly. Seat positions were recorded (as adjustment notches Back and Down from "FULL UP AND FORWARD"). Pilots will need to place their eyes inside the HUD EMB to accurately see 100% of the HUD symbology. If pilots need to move from their normal seat position, accommodation problems (eg. clearance and/or reach problems) could occur (Hudson, Zehner, Harbour, & Whitehead, 2011**.** To maximize the ecological validity of this study, all data collection was accomplished in an actual aircraft cockpit on the airport tarmac where actual flight operations occur (Figures 1 and 2). To avoid disrupting the subject's normal habit patterns in the cockpit, experimenters did not direct the methods that a pilot would use to control the aircraft. Consequently, this study is a quasi-experimental design.

Figure 1. FARO Arm mounted in co-pilot position after seat removal.

Figure 2. Installed FARO Arm for cockpit and subject digitization.

Participants and Procedures

Participants. Nineteen participants were recruited from the Great Lakes region and ranged in ages from 20 to 60. Seven of the participants had flying experience and four were qualified in the aircraft. Participants occupied the pilot position using both the Heads Down Display (HDD) and HUD for ground testing. The test plan was reviewed and approved by the Aeronautical Systems Center's Technical Review Board. As the study was part of normal duties, no additional compensation was provided.

Tasks. Each test case was outfitted with and without operational gear. Each subject's accommodation was evaluated in the cockpit: internal field of view (IFOV), external over the nose (OTN) vision, reach to controls, overhead clearance, and egress (where required). HUD FOV (HFOV) was also evaluated.

Equipment. The military aircraft had an L-3 Communications HDD and a Rockwell Collins Flight Guidance Systems HUD. The digitized data were collected using a FARO Arm system (up to 0.0007" accuracy), which included a cockpit mountable 3-D laser data collection system.

Procedures. Reach measurements were taken on subjects reaching to the landing gear handle, and the upper, middle, and lower central main instrument and computer panel switches. This was done with and without straining (shoulder, arm, and leg muscles) against a locked shoulder restraint system. Control authority (Yoke, Rudders, and Throttles) was also measured and assessed. Nose wheel steering was not tested. Subject seat positions associated with optimum HUD EMB position, were compared to the mapped accommodation results (Rudder authority, Yoke Pitch and Roll Clearance, etc.) which were recorded throughout the seat position range. A Pass / Marginal /Fail, (coded GREEN / YELLOW / RED) were assigned on these issues.

Definitions. **Cockpit Design Eye Point (Cockpit DEP).** Spatial location and pilot position where the pilot should sit in order to operate the aircraft for optimal visibility both inside and outside, have optimal aircraft controllability and cockpit reach, and proper outside visibility to Take-Off, Fly, and Land the aircraft safely in the way it was designed, while accommodating the pilot population. Theoretically, the adjustment mechanisms for the seat and the rudders would offer accommodation for the variation in pilot body size and proportion. A set of design eye spheres, two per side, (Figure 3.) were located adjacent to the clock, above the glare shield, to aid the pilots in acquiring an eye position on Cockpit DEP.

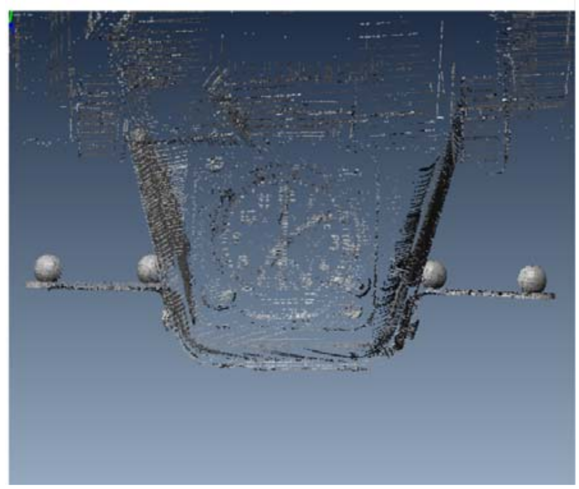


Figure 3. Digitized Cockpit "Design Eye Spheres."

HUD Eye Motion Box (HUD EMB). A three-dimensional envelope within which the pilot's eyes need to be in order to accurately see 100% of the symbology. **Total Field of View (TFOV).** The spatial angle in which all of the the symbology can be displayed / viewed measured laterally and vertically. **GREEN** – No control interference. **YELLOW** – Some control interference may occur but pilot is able to move leg/s out of the way and or is still able to move controls even if they are impacting gear. However, it complicates control of the aircraft and may delay required inputs and may still have lost up to last 5% of control authority. **RED** – Significant control interference occurs and movement of controls are impeded (limits control authority). Approximately up to last 30% of pitch aft control authority is lost and up to 30% of bank authority is lost (subject dependent).

RESULTS

The pilot and copilot Cockpit DEP locations were reverse engineered from the design eye spheres using the FARO Arm, and were located relative to the rest of the digitized geometry using an arbitrary Cartesian coordinate system (x, y, z). This was necessary because: 1) the cockpit dimensional drawings were not available, and 2) the geometric justification for HUD system placement was not known. The TFOV and HUD EMB Center were mapped utilizing actual subject pupil locations and the FARO Arm (Figure 4).

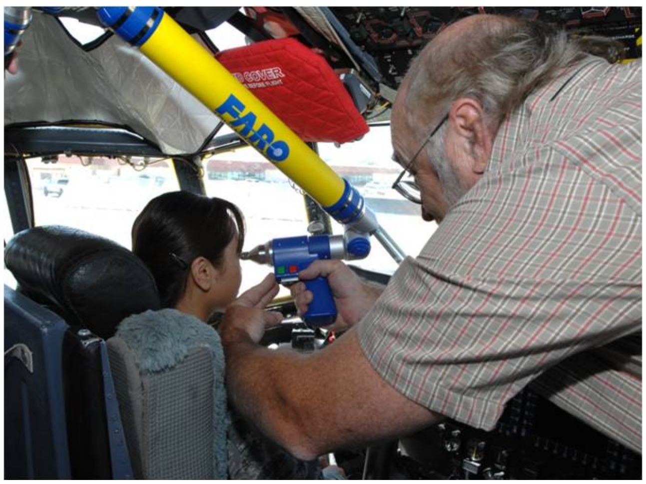


Figure 4. Digitizing Subject Eye Location.

All subject specific location differences between mid-pupil at "Perceived DEP" and at HUD Eye Motion Box center was geometrically calculated. These optical differences could be virtually used to translate a subject into a seat position where the subject would sit to place his or her Mid-Pupil at the HUD Eye Motion Box Center, where other accommodation issues could be addressed.

It was found that for the installed HUD, the Cockpit DEP & HUD EMB did not match; represented by an average vertical distance difference of ~2 inches (range of 1.6 to 3.25 inches), (Figures 5 and 6). Therefore, this reduced visible HUD symbology resulted in a minimum loss of 25% to as much as a 100% loss depending on pilot perception of DEP.

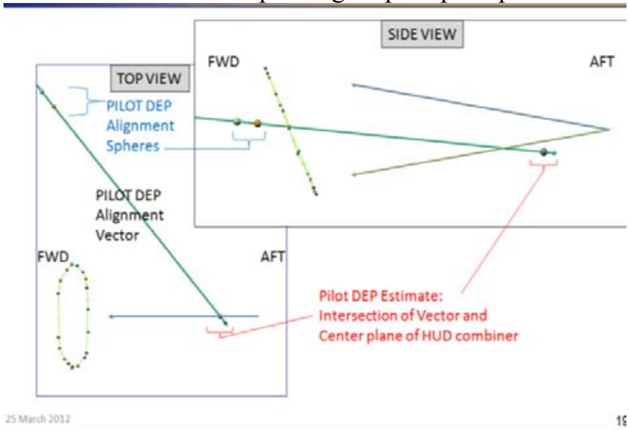


Figure 5. Defining the Cockpit Design Eye Point for Pilot side using the original Design Eye Spheres.

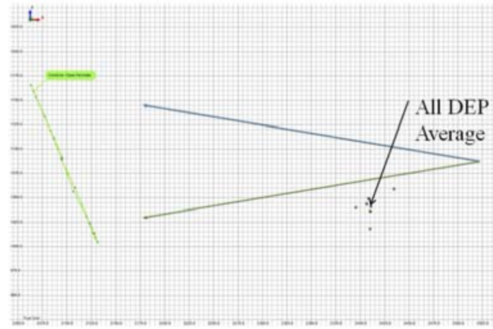


Figure 6. Average of Cockpit DEP locations calculated from six different pilot stations and three different aircraft superimposed with HUD EMB. (Pilot side view with Combiner and Aircraft Nose to left.)

The HUD symbology (Figure 7) loss was measured and mapped when mid-pupil was outside of HUD EMB and TFOV. The percentage of symbology lost (½" out, 1" out, etc.) is mapped below in Figure 8. Although pilots are known to do it, "Head Tilting" is not considered when writing specifications or quantifying accommodation.

Figure 7. View of combiner when sitting in CDEP with a 25% loss of visible symbology (most common). Some subjects experienced a total loss of visible symbology.

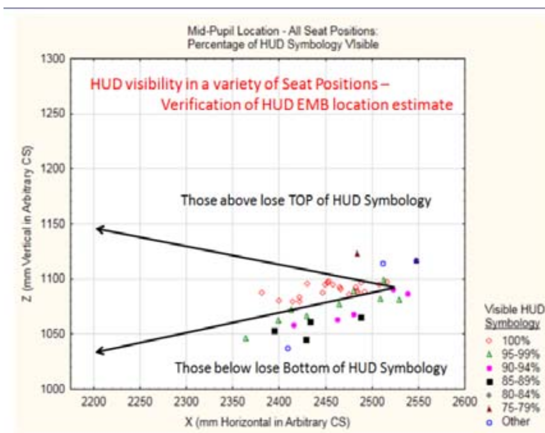


Figure 8. Percentage of HUD symbology visible outside of HUD EMB. (Pilot side view with HUD Combiner on left.)

To illustrate using subject photographs, below in Figure 9, a subject of average pilot eye height, is seated so his eye is at Cockpit DEP. Figure 10, indicates the extent of needed head tilt and neck stretch to place his eye into a position to use the HUD symbology.

Figure 9. Subject placed at Cockpit DEP. Eye is below and aft of HUD EMB center.

Figure 10. Subject placed at Cockpit DEP and then asked to move his head in order to see the HUD symobology.

DISCUSSION

The mismatch of the prototype HUD EMB and the Cockpit DEP (i.e. HUD EMB Center relatively ~2" higher and ~1" forward) directly impacts accommodation. A raised seat to place a pilot eye in the HUD EMB will create yoke interference (Figure 11), as well as longer reaches to rudders and controls downward. Conversely, if the pilot remains at Cockpit DEP, the loss of visible HUD symbology results in at least 25% (above, Figure 5) and as much as a 100%, depending on the pilot's eye positioning, which is based on their perception of the Design Eye Sphere visual line.

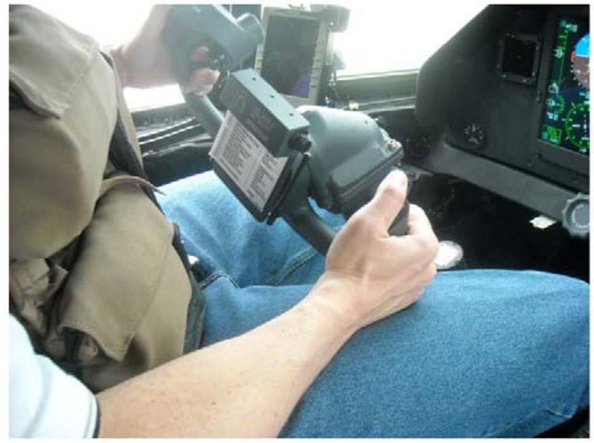


Figure 11. Yoke control interference was SEVERE when subject eye was placed in HUD Eye Box - 18 out of 19 subjects FAILED (with Survival Vest on).

This optical point discrepancy restricts the pilot population while potentially increasing Work Load (WL) and potentially decreasing Situational Awareness (SA) (Figures 12 and 13). Ultimately, the mismatch reduces mission capability, effectiveness, and safety.

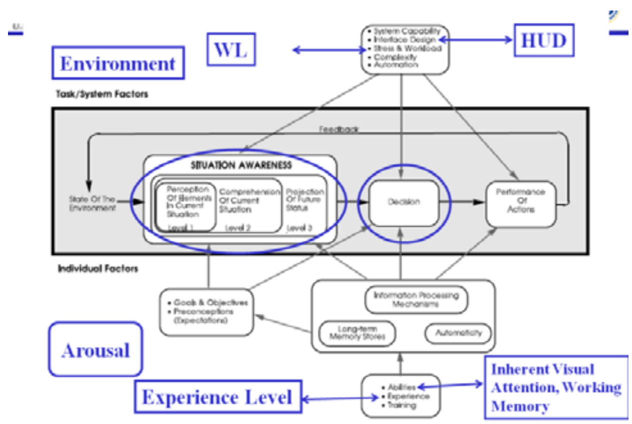


Figure 12. Theoretical Model of SA (adapted from Endsley, 1995b). Items in blue are additional considerations added by Harbour.

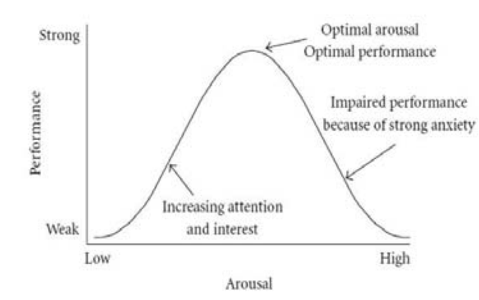


Figure 13. Effect of Arousal on Performance. Yerkes & Dodson, 1908. The cognitive neuroscience factors forming SA and contributing to WL, which could be affected by the spatial location of the HUD.

Recommendation I: Install HUD combiner and projector spacers in order to approximately match CDEP and HEMB. This was implemented.

Recommendation II: Given enough time, develop a new HUD prototype that given more data and a larger TFOV will exactly match CDEP and HEMB. This is being implemented.

The authors note that their own observations of subjects in the pilot position while seated at Cockpit DEP would need to strain their neck in order to see HUD symbology and then **also** have to look down approximately 40 degrees in order to see the HDD PFD, in order to cross-check what was seen in the HUD. This is bound to increase physical fatigue, and mental workload, in addition, creating potential issues due to the pilot's attention switching back and forth between the HUD and the HDD. Not all of the PFI on the HUD is visible due to the HUD EMB and Cockpit DEP mismatch, creating a time disruption in the interpretation of attitude information or other PFI and the effect of such transitions on pilot SA is unknown.

Impact to Accommodation:

The mismatched eye positions of HUD and Cockpit DEP not only potentially increases Work Load (WL), and potentially decreases Situation Awareness (SA), but it also reduces mission capability & effectiveness. In Figure 14, below, a qualitative assessment, based on our subject anthropometry, compares control authority for Yoke and Rudder while sitting at the current HUD EMB to that of the Cockpit DEP while the required armor vest is worn, both with and without the survival vest.

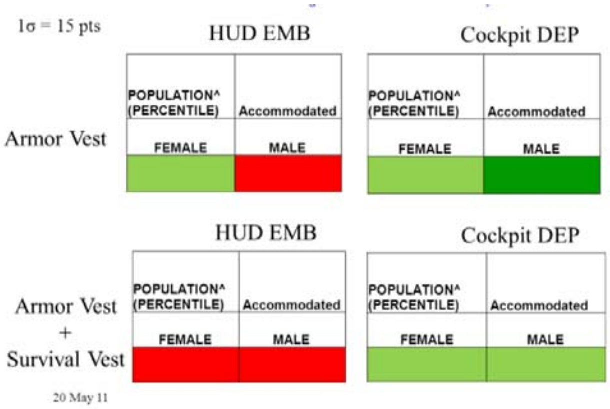


Figure 14. Adverse impact to accommodation (Yoke and Rudder authority) as a function **of a more forward and higher** HUD EMB as compared to the CDEP (Red is "no-go" and Green is "go"). Results for subjects wearing Armor Vest with and without Survival Vest are shown.

Lowering the HUD system in order to allow HUD EMB to match Cockpit DEP (Figure 15) would yield the best case scenario for mission success and safety, and best aircraft controllability.

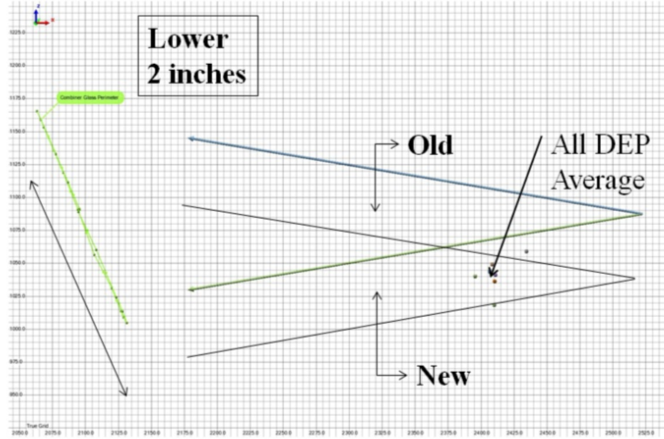


Figure 15. Entire HUD system needs to be lowered 2 inches to match HEMB and CDEP.

Figure 16. Pilot-Author with eye at Cockpit DEP and below HUD EMB.

CONCLUSIONS

The results presented here indicate that a method should be established to ensure that the HUD is ideally spatially located in the cockpit to accommodate pilots and potentially reduce pilot WL and increase SA, which is the intent of the HUD. This study presents such a method even when cockpit drawings and HUD system optical characteristics are not known by sampling the pilot population and digitizing pupil locations, and the cockpit and HUD geometry. More research in this area needs to be accomplished, blending ergonomics, optics, and cognitive neuroscience in the actual aircraft in-flight. Neuroergonomics is a new field that integrates research between psychology, cognitive neuroscience, engineering, and ergonomics (Parasuraman, Christensen, & Grafton, 2011). The effects of varying spatial locations of information displays (ID) in addition to individual differences in visual perception and attention coupled with the effects on pilot WL and SA should be researched next (Harbour, Christensen, Estepp, & Gray, in press; Tsang & Vidulich, 2006; Wickens & McCarley, 2008).

REFERENCES

Air Force Flight Standards Agency. (2007). *White paper: Primary flight reference endorsement process.* Washington, DC: Author.

Endsley, M.R. (1995b). Toward a theory of situation awareness in dynamic systems. Human Factors 37(1), 32–64.

Flight Safety Foundation. (2009). *Head-up guidance system technology – A clear path to increasing flight safety.* November 2009 Special Report. Alexandria, VA: FSF

Harbour, S.D., Christensen J.C., Estepp J.R., and Gray, T.M. (in press). *An integrated neuroergonomic assessment of in-flight pilot workload.* 56th Annual Meeting of the Human Factors and Ergonomics Society, Boston, MA (Oct 22-26, 2012).

Hudson, J.A., Zehner G.F., Harbour S., Whitehead, C.R. (2011). *Design eye point vs. hud eye Box vs. pilot eye position: A 3d location comparison.* In the Proceedings of the 49th Annual Symposium of SAFE, Reno, NV (Oct 24-26, 2011).

Parasuraman, R., Christensen, J., Grafton, S. (2011). *Neuroergonomics: The brain in action and at work.* . NeuroImage (impact factor: 5.74). 08/2011; 59(1):1-3. DOI: 10.1016/j.neuroimage.2011.08.011. Fairfax, VA

Tsang, P.S., and Vidulich, M.A. (2006). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (3rd ed., Chap. 9). New York: Wiley.

Wickens, C. D., & McCarley, J. S. (2008). *Applied attention theory*. Boca Raton, FL: Taylor and Francis Group.

Yerkes RM, Dodson JD (1908). "The relation of strength of stimulus to rapidity of habit-formation". *Journal of Comparative Neurology and Psychology* **18**: 459–482.

COMMUNICATION SEQUENCES IN CONTROLLER PILOT COMMUNICATIONS

Jonathan Histon
University of Waterloo
Waterloo, ON, Canada

Esa Rantanen and Cecilia Ovesdotter Alm
Rochester Institute of Technology
Rochester, NY, USA

Temporal patterns in the content of controller-pilot communications can reflect strategies and techniques used by controllers. Introducing data-based communication systems may profoundly impact these strategies and techniques. To establish a baseline of existing patterns in controller-pilot voice communications, analysis is presented of the content of air-ground communications between pilots and controllers from North American operations. Results are presented of the relative frequency of common sequences of communication events and how these vary in response to changes in weather conditions. Methodological challenges in sequence analysis as well as implications of the results for understanding effect of changes in communication technology on controller training and complexity management in future ATC environments are presented.

Introduction

This research examines the current, typical actions controllers perform on aircraft to better understand the strategies that controllers use in existing voice-based communications environment. Controller-pilot communications represent the primary (but not only) mechanism available to the controller to affect the evolution of the system being controlled. In particular, controller-pilot communications offer a robust indirect means to examine numerous aspects of controller performance and strategies that vary under different operational conditions, such as controller workload or weather conditions. In this paper we investigate sequences of transmissions from controllers to pilots. These sequences provide insight into the degree of standardization of the existing task. Examination of differences in the sequences under different conditions provides insight into how controller strategies vary with conditions. The primary focus of this paper is on documenting the analysis process. We also report initial results comparing sequences used in different weather conditions.

Previous Work on Communications Analysis

It is hardly surprising that controller-pilot communications have been studied extensively over the past several decades. Air traffic controllers' work is primarily cognitive and therefore difficult to observe. Until very recently, voice communications between controllers and pilots, and between controllers responsible for different airspaces, have provided the only observable indicators of controllers' decision processes, strategies, and workload. Because voice has been the only link between controllers and pilots, such communications have been used as an indirect measure of myriad variables of interest. Very briefly, voice communications have been analyzed in different ways to examine covert constructs such as controller workload (e.g., Manning et al., 2002), controller situation awareness (e.g., Metzger & Parasuraman, 2006; Gil et al., 2008), controller performance through communication errors (e.g., Prinzo et al., 2007; Cardosi, 1993), trust and preference for communication modes (e.g., Stedmon et al., 2007; Metzger & Parasuraman, 2006; Sharples et al., 2007), controller strategies (e.g., Histon, 2008; Filho, 2012), and language proficiency (e.g., Prinzo et al., 2008).

Controller workload is one of the most critical aspects of controller performance. Numerous studies have shown workload and communications having a strong and robust interrelationship, making communication activity a preferred nonintrusive measure of workload (McGann et al., 1998; Manning et al., 2002). Specific communication variables that can be measured from ATC voice recordings or transcripts include duration of and frequency of verbal communication events (e.g. Hurst & Rose, 1978; Manning et al., 2002; Porterfield, 1997). Rantanen, Naseri, and Neogi (2007) analyzed diverse operational data from different en route center sectors and showed strong correlations between metrics such as dynamic density (Laudeman, Shelden, Branstrom, & Brasil, 1998) and number of conflicts and controller communication time. In experimental settings, manipulating communication load affects both subjective ratings and objective measures of workload (Casali & Wierwille, 1983). A more detailed review of past work on ATC communication analysis is provided in Histon, Rantanen, and Alm (2012).

Continuing Importance of Communications Analysis

With the advent of the NextGen ATC modernization the traditional means ATC has been done, that is, through voice communications between controllers and pilots, will fundamentally change. Although this change is inevitable and in many cases for the better, for voice communication is prone to errors and a clumsy way to accomplish the goals of ATC, it is important to carefully monitor also the unintended consequences of NextGen. For example, the loss of so-called "party-line" information (Hansman, et al., 1998) has raised concerns about the ability of flight crews to maintain situation awareness (Signore & Hong, 2000). Although this loss in awareness may be partly offset by the ability to retain and review controller-pilot datalink (CPDLC) messages (Metzger & Parasaruman, 2006) and the less disruptive nature of CPDLC where the text from a command can be stored and recalled later to ease a pilot's working memory and prevent task disruption. Yet, the use of CPDLC will in all likelihood significantly alter conflict resolution strategies and representations of the traffic situation shared amongst controllers (Kapp & Celine, 2006).

To systematically evaluate the changes in controller strategies, situation awareness, and workload, which remain all-important components of controller performance also under fully implemented NextGen, it is critical to have a valid and established baseline against which any changes—for better of worse—can be measured. The analyses presented in this paper focus on sequences in the content of the communications and the implementation of controller commands. As far as we know, these particular aspects of controller-pilot communication have not been systematically researched before.

Method

To document the actions controllers currently perform, nearly 90 hours of voice-based communication data were collected and coded. The following is an abbreviated description of the analysis process; a complete description can be found in Histon (2008). Recordings of two-way controller-pilot communications were obtained from two internet websites: www.atcmonitor.com and www.liveatc.net. These websites archive and stream live controller-pilot radio communications using private radio scanners. Observations were collected for seven en route sectors in the US. The sectors represent a range of types of operation (en route arrival and departure sectors, and sectors containing mostly overflights) and cover a range of different altitude strata. Data were roughly categorized into "good weather" and "poor weather" conditions by historical weather radar images. For each time period the data were collected, an analyst classified the time period as "good weather" or "poor weather" based on the presence of widespread convective returns in the general geographic area of each sector.

The time, aircraft addressed, and content of each transmission from the controller were determined. The coding scheme focused on controller–pilot communications; with the exception of pilots announcing their presence on frequency ("check-in" transmissions), pilot-controller communications were not coded. To categorize the content for analysis, each transmission by a controller was reduced to elemental communication events, or the smallest decomposition of parts of a transmission that would retain meaning to the recipient. For example, the transmission "Turn left twenty degrees for spacing" was parsed into the elements of "turn left twenty degrees" and "for spacing." Each elemental event was stored as a separate entry in the database.

Elemental communication events were grouped into seven "Categories": (1) Command, (2) instructions, (3) gathering information, (4) giving information, (5) handoff, (6) other, and (7) unknown. Each category was further subdivided into individual "Types" of events. For example, "Commands" were defined as elemental communication events that modified an aircraft's clearance either by requiring or permitting a modification to the aircraft's trajectory. The results of the coding were collected and archived in a SQL Server database.

A time-ordered description was developed of the elemental communication events for each flight in the data set (for example see the left column in Table 1). The sequence as a whole was then used for subsequent analysis and reporting. Due to the way data were stored in the database, sequences could be generated reflecting three different levels of abstraction of the communication events: (1) "Specific" sequences, including all details of each element of each transmission, (2) "Types of" sequences, retaining only the "Type of" each element of each transmission, and (3) "Categories of" sequences, retaining only the "Category of" each element of each transmission. An example of each type of sequence is shown in Table 1. In addition, filters were developed allowing analysts to restrict which communication events were included in a sequence. This was used to focus, for example, just on the commands employed by controllers. It also allowed analysts to eliminate superfluous parts of the data collected; for the purposes of the analysis presented here, "Roger/Acknowledgement" communication events, which were originally collected, were filtered out of the data set.

Table 1.
Examples of Same Sequence at Different Levels of Abstraction (Data From Sector C).

Specific Sequence		"Type of" Sequence		"Category of" Sequence	
Sequence	% of Flights	Sequence	% of Flights	Sequence	% of Flights
•Gave Altimeter Setting •Cross <MULRR> at <100> (FL / 100 Feet) •Cross <MULRR> at <250> Knots •Checkout to <EVANSVILLE APPROACH> (126.1)	4.0%	•Altimeter Setting •Cross <Fix> at <X> Feet •Cross <Fix> at <X> Knots •Checkout	4.6%	•Providing Information •Command •Command •Handoff	7.9%

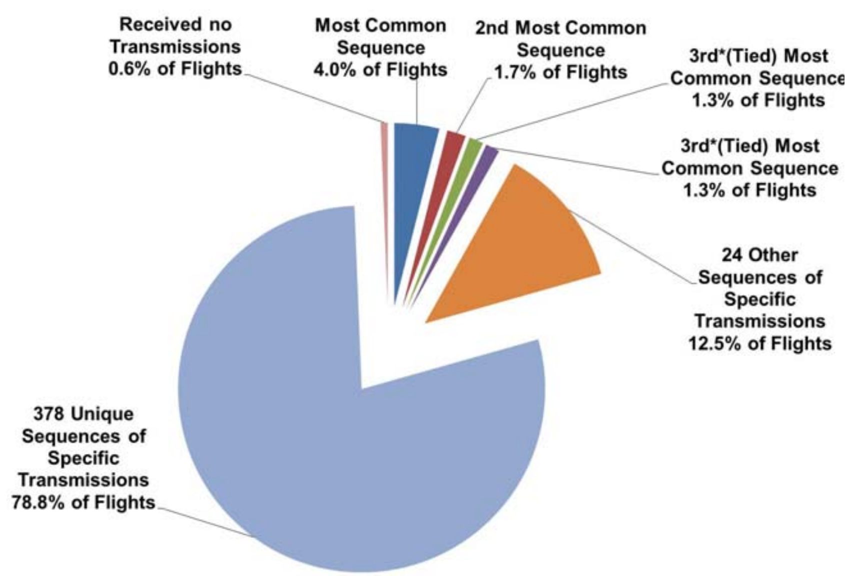
The results discussed below include consideration of unique sequences; a unique sequence is a sequence that one and only one aircraft received within the observation time for each sector. The results below report the total number of such sequences, rather than presenting each individual, unique, sequence.

Results

To demonstrate the effect of varying the level of abstraction on the frequency of sequences, Sector C is used as an illustrative case study. Figure 1shows the top 3 (and ties) most frequently occurring sequences when analyzing all transmissions at a "Specific" level of detail. The most common sequence reflects a standard pattern common to sectors that contain descending aircraft: an altimeter setting followed by a crossing restriction. In over 15.5 hours of data and nearly 500 flights, 75% of the "specific" sequences were completely unique to each aircraft.

The pattern of having a large number of unique "specific" sequences was observed across all sectors. More than half of all flights in every sector were unique in terms of the communication events generated by the controllers. This has important implications for both training and data-based operational concepts. A clear takeaway is that training is not about simply generating pre-determined scripts of actions; as much as observational evidence suggests there are regular recurring patterns in controller communications (Histon, 2008), in operational practice the specific communication events show significant variation in content. In addition, the high percentage of unique sequences also indicates that it is unlikely that standardized command sequences could be easily defined for a sector. Such standardized sequences could be one way of taking advantage of the ability of data-based communications.

To investigate whether similar patterns were found when the communication events are considered at a more abstract level, the same analysis was repeated for sequences generated using only the "Categories of" description of



Sequence	% of Flights
•Gave Altimeter Setting •Cross <MULRR> at <100> (FL / 100 Feet) •Cross <MULRR> at <250> Knots •Checkout to <EVANSVILLE APPROACH> (126.1)	4.0%
•Gave Altimeter Setting •Descend and Maintain <90> (FL / 100 Feet) •Checkout to <ZNY> (124.1)	1.7%
•Checkout to <ZNY> (132.5)	1.3%
•Direct to LRP (LANCASTER) •Cross <30 W OF LANCASTER> at <170> (FL / 100 Feet) •Gave Altimeter Setting •Checkout to <ZNY> (128)	1.3%

Figure 1. *"Specific Transmissions" sequences and % of flights observed for Sector C. Table shows details of top 3 (and ties) most common sequences.*

each communication event. The number of unique sequences was cut in half, while the proportion of flights receiving the most common sequences increased slightly (*Figure 2*).

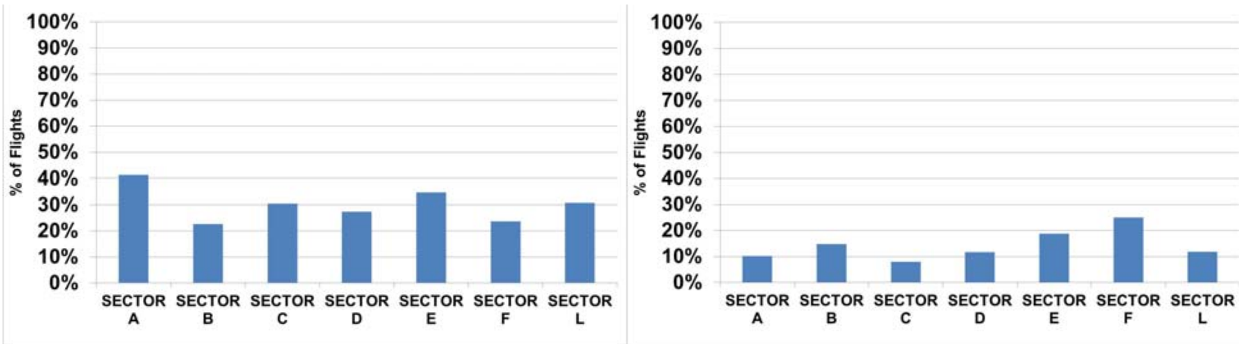


Figure 2. *Left, percent of flights receiving unique "Categories of" Transmissions" sequence for each sector. Right, percent of flights receiving the most common "Categories of" Transmissions" sequence in each sector.*

The analysis above took into account all of the transmissions in the dataset. To focus solely on actions taken by controllers that change an aircraft's trajectory, filters were created restricting the dataset to only those transmission events that are part of the "Command" category. Figure 3 (left) shows that the percentage of unique sequences of "Types of Commands" is relatively small but with some substantial sector-to-sector variation. The percentage of flights that receive the most common sequence is consistently approximately 10% across sectors (Figure 3, right).

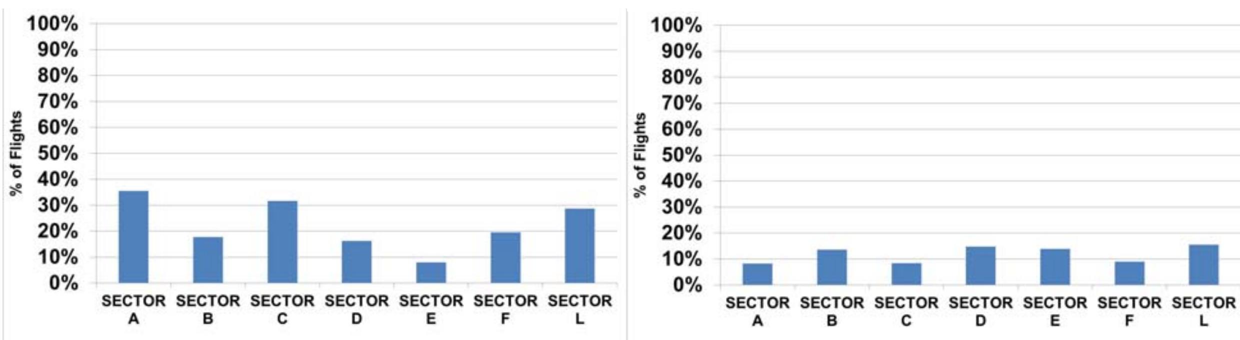


Figure 3. *Left, percent of flights receiving unique "Type of" commands sequence for each sector. Right, percent of flights receiving most common "Type of" commands sequence in each sector.*

These results illustrate how sequences of communication events can be analyzed at different levels of abstractions, and how the relative frequency of sequences varies across en route sectors. Of interest is also considering how sequences change under different operating conditions. Figure 4 shows the percentage of flights receiving unique sequences of commands generally increases under "Poor" weather conditions; this is consistent with the expectation that the presence of convection, poor rides, or other factors associated with poor weather would generate more pilot requests and more need for interventions in the system. There is also a general corresponding decrease in the percentage of flights that did not receive any commands at all.

Discussion

This paper is an initial exploration of the use of analysis of sequences of events. As described above, analysis can be conducted at different levels of abstraction of the communication events, and with different filters applied to support targeted analysis. Further work will address a number of opportunities to refine the techniques being developed and expand their use in controller strategy identification. In particular, the database has been developed in a way that promotes flexibility and the ability to incorporate alternate forms of analysis. For example, additional levels of abstraction could be introduced to refine analyses of commands; all altitude commands, or all speed commands can be grouped together providing the ability to discern patterns in which forms of maneuver are used and how this varies with conditions such as traffic load and/or weather conditions. There are also opportunities for integrating the use of computational linguistics modeling techniques for natural language data.

In addition, further analysis could incorporate the work of Filho (2012) in identifying subsets of traffic within each sector and examining the consistency of sequences for each grouping. While access to radar track data is not generally available, the sector an aircraft is handed off to is available for most of the data set and can be used as a crude measure for establishing groups. Alternate formulations of the underlying taxonomy could also be (relatively) easily introduced into the database and used for analysis. Distinguishing between different subtypes of en route sectors, for example distinguishing high level overflight sectors from low level arrival and departure sectors would allow for further comparisons and investigation of control action differences.

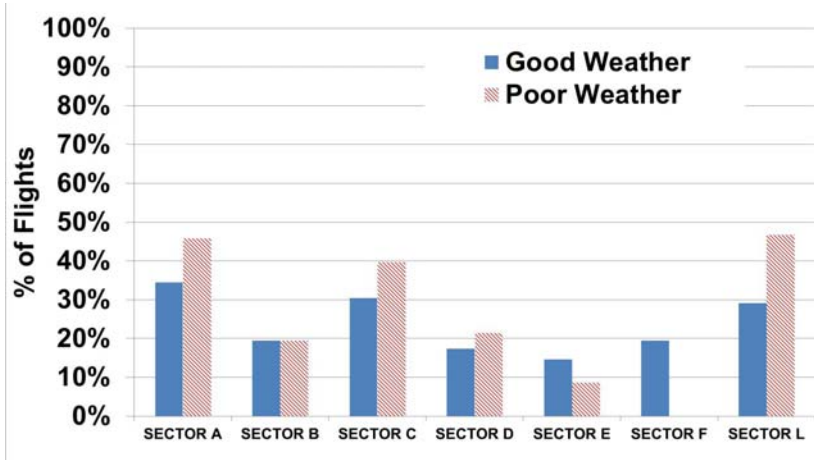


Figure 4. *Percent of flights receiving unique "Type of Command" sequences in "Good" and "Poor" weather for each sector.*

One of the challenges with the analysis of unique sequences presented above is that it is sensitive to the amount of data collected in each sector. The threshold for a unique sequence was always one aircraft; however, the likelihood of another aircraft sharing the same sequence of communication events is dependent on how long a time period is being analyzed. The longer the time period, the greater the chance of another aircraft receiving the same sequence. A more appropriate metric would be to consider "unique" sequences to be those that are received by less than a fixed percentage of aircraft (e.g. less than 0.5%). Using this definition, the proportion of unique sequences would not be affected by changes in the duration of data collected.

Finally, there are several methodological challenges to be addressed. A limitation of the current method is that it relies on analysts being able to hear and record the call sign for each flight. This can be difficult to maintain consistency of, particularly when a flight is present in two consecutive audio files; each file may be listened to by a different analyst, and establishing that the same call sign applies to flights recorded with different call signs is an ongoing challenge. While the data presented in this paper has been subject to quality control and validation tests, further work is needed to develop tools for making it easier to identify and correct mismatches in the data set.

Conclusion

In this paper we have reported novel means of analyzing controller-pilot communications. The primary purpose of the research was to develop methods for investigation of current patterns of communication that can be used to infer controllers' strategies, situation awareness, and workload. Communication sequences are very important in the current voice communication environment as standard sequences help reduce both controller and pilot workload and errors. Deviations from standard sequences may indicate nonroutine situations or high workload or degraded situation awareness. The latter inferences would have to be correlated with and corroborated by other measures, however. Our primary purpose with this research has been to establish a baseline of controller work to which changes brought about implementation of NextGen can be compared. Because of the envisioned extensive use of CPDLC under mature NextGen, it seems very important to have a thorough understanding of current ATC practices through objective measures derived from operational data.

References

Cardosi, K. M. (1993). Time required for transmission of time-critical air traffic control messages in an en route environment. *International Journal of Aviation Psychology, 3*(4), 303–313. doi:10.1207/s15327108ijap0304_4

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.

Filho, E. de A. (2012). *Analysis of airspace traffic structure and air traffic control techniques*. Unpublished Master's Dissertation, Massachusetts Institute of Technology.

Gil, F. O., Vismari, L. F. & Camargo Junior, J. B. (2008). Analysis of the CPDLC real time characteristics and the Mode S data link capacity. *Anais do VII SITRAER Anais... Rio de Janerio: Universidade Federal do Rio de Janeiro, v.1*, p. 92-99.

Hansman, R. J., Endsley, M., Farley, T., Vigeant-Langlois, L., & Amonlirdviman, K. (1998). The effect of shared information on pilot/controller situation awareness and re-route negotiation. *Proceedings of the 2nd FAA/Eurocontrol International Air Traffic Management R&D Seminar (ATM 98)*, Washington DC: FAA.

Histon, J. M. (2008). *Mitigating complexity in air traffic control: The role of structure-based abstractions*. Unpublished doctoral dissertation, Massachusetts Institute of Technology. Available from http://hdl.handle.net/1721.1/42006

Histon, J., Rantanen, E., & Alm, C. O. (2012). Air traffic control voice data analysis to validate NextGen procedures. In S. Landry (Ed.), *Advances in Human Aspects of Aviation (Advances in Human Factors and Ergonomics Series)* (pp. 129-138). Boca Raton, FL: CRC Press.

Hurst, M. W., & Rose, R. M. (1978). Objective workload and behavioural response in airport radar control rooms. *Ergonomics, 21*(7), 559-565.

Kapp, V., & Celine, C. (2006). Designing an ATC CPDLC environment as a common information space. *Proceedings of 2006 IEEE/AIAA 25TH Digital Avionics Systems Conference*, Portsmouth, OR.

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. 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.

McGann, A., Morrow, D., Rodvold, M., & Macintosh, M. A. (1998). Mixed-media communication on the flight deck: A comparison of voice, data link, and mixed ATC environments. *International Journal of Aviation Psychology 8*, 137–156.

Metzger, U. & Parasuraman, R. (2006). Effects of automated conflict cuing and traffic density on air traffic controller performance and visual attention in a datalink environment. *International Journal of Aviation Psychology 16*(4): 343-62.

Porterfield, D. H. (1997). Evaluating controller communication time as a measure of workload. *International Journal of Aviation Psychology, 7*(2), 171-182.

Prinzo, O. V., Hendrix, A., Hendrix, R., (2007). The computation of communication complexity in air traffic control messages. *Proceedings of the 7th FAA/Eurocontrol ATM R&D Seminar*. Barcelona, Spain.

Prinzo, O. V., Hendrix, A. M., & Hendrix, R. (2008). *Pilot English language proficiency and the prevalence of communication problems at five U.S. air route traffic control centers*. (DOT/FAA/AM-08/21). Washington, DC: FAA.

Rantanen, E. M., Naseri, A., & Neogi, N. (2007). Evaluation of airspace complexity and dynamic density metrics derived from operational data. *Air Traffic Control Quarterly, 15*(1), 65-88.

Sharples, S., Stedmon, A., Cox, G., Nicholls, A., Shuttleworth, T. & Wilson, J (2007). Flightdeck and air traffic control collaboration evaluation (FACE): Evaluating aviation communication in the laboratory and field. *Applied Ergonomics 38*(4), 399-407.

Signore, T. L., & Hong, Y. (2000). Party-line communications in a data link environment. *In 19th Digital Avionics Systems Conference, Vol. 1* (pp. 2E4–1–2E4/8). IEEE.

Stedmon, A. W., Sharples, S., Littlewood, R., Cox, G., Patel, H., & Wilson, J. R. (2007). Datalink in air traffic management: Human factors issues in communications. *Applied Ergonomics, 38*(4), 473–480. doi:10.1016/j.apergo.2007.01.013

Recommendations Supporting Development of Flight Deck DataComm Text and Graphic Display Evaluation Guidance

Jennie J. Gallimore, Ph.D.
Ricardo D. Munoz, Stephen B. Kiss, Chang-Geun Oh, Timothy Crory, Randall Green,
Clark Shingledecker, Ph.D., Pamela S. Tsang, Ph.D.
Wright State University
Dayton, OH 45435

In the Next Generation Air Transportation System (NexGen), voice communications will become less frequent, and most communication will occur via data communications -- uplink messages (UM) (to pilot) and downlink messages (DM) and requests (to ATC). Clearances may include simple one-element clearances such as CLIMB TO [altitude] or complex clearances created by concatenating messages to create flight trajectories that include ATC-authorized route segments, altitudes, and at least one required time of arrival (RTA). Due to the complexity of clearances, aircraft and flight deck equipment manufacturers may seek approval for new and modified flight deck displays to more clearly depict clearances to the flight crew, likely using text and graphics. This research evaluated text and hybrid text and graphic concepts to develop human factors (HF) recommendations for specialists who participate in certification of new and modified flight deck DataComm displays, and as a potential update to AC 20-140, Guidelines for Design Approval of Aircraft Data Link Communication Systems Supporting Air Traffic Services (ATS).

Data communications (DataComm) is one of the key technologies supporting the transition to NextGen. DataComm refers to the communication between air traffic controllers (ATCs) and pilots which will change from voice clearances to satellite datalink communications. DataComm is a transformational program that is critical to the success of NextGen operations. It will provide infrastructure supporting other NextGen programs and operational improvements, and enable efficiencies not possible using air/ground voice communications alone. Because DataComm is a key enabling technology that significantly affects human performance, human factors experts have anticipated potential implementation challenges (Cardosi, Lennertz and Donahoe (2010).

One challenge for the flight crew will be understanding Trajectory Based Operations (TBO) clearances. Textual clearance displays that provide complex 4D trajectory information may be difficult for pilots to interpret in a timely and efficient manner without error. TBO will require spatial understanding of the location of the aircraft with respect to location in 3D space as well as time. Presenting spatial information to pilots via text alone requires pilots to perform a mental transformation that could slow down the understanding of the messages and lead to interpretation errors. This research investigated use of alternative flight deck displays with graphics, hybrid text and graphics, and other formats that could be integrated with existing navigation displays (NDs) or new DataComm displays to enable pilots to more easily identify, understand, and quickly respond to air traffic clearances and instructions. Alternative displays may also better support negotiation of clearances. The purpose of this research was to evaluate text and hybrid graphics and text concepts in order to support the FAA's Aircraft Certification Service need for regulatory guidance to evaluate alternative flight deck displays, and to make recommendations for minimum requirements for system characteristics and display of air traffic clearances on the flight deck.

Background

While studies that examine the effects of presenting information graphically to the pilots in the cockpit on pilot-controller communications are beginning to emerge (e.g., Prinzo, 2003; Wickens et al., 2003), there is a paucity of studies on graphical display of clearance instructions. One early study by Hahn and Hansman (1992) was focused on the relationship of situational awareness to automated Flight Management System (FMS) programming of data linked clearances and the readback of ATC clearances. Situational awareness was tested by issuing nominally unacceptable ATC clearances and measuring whether the error was detected by the subject pilots. The study also varied the mode of clearance delivery: verbal, textual, and graphical. Results showed that graphic depiction of data link routing information received from the controller and embedded in the electronic map display imposed lower workload than either text or spoken representation of the same spatial information. These researchers pointed out that because textual and graphical modes of clearance delivery offered different advantages for processing, a combination of these modes of delivery in a data link presentation might be advantageous.

A research study was conducted that evaluated 39 different ATC clearances from the RTCA SC-214 / EUROCAE WG-78 Standards for Air Traffic Data Communication Services, referred to here as the SC-214 message set under both text and hybrid graphic and text format conditions. The 39 clearances included 1, 2, 3, 4, 5, 6 or 9 element clearances. A single UM with two elements is AT [Position] CLIMB TO [level], where the element is the variable in the clearance that changes. The one nine-element clearance was composed of one UM (UM339) concatenated together three times to give the pilot a new trajectory.

Method

Experimental Design

The experimental design was between-subjects with one independent variable, presentation FORMAT, with five levels. Number of elements is considered a control variable. It was not feasible to create a factorial design with number of elements as a variable because the specific concatenated UMs were different depending on the number of elements. However, this variable allows us to analyze the data at each level of clearance, and to examine performance trends as the number of elements increase.

Graphic Formats

The five formats included the baseline condition of TEXT only, Graphics + Text, Graphic + Text with updated SC214 UMs, Graphics + Text + altitude situation display (ASD) and Graphics +Integrated Text +ASD. Each are brefily described below

TEXT. The text condition included the presentation of a navigation display (ND) with clearances presented in text to the right of the ND. The ND included the current flight path.

Graphics + Text (G+T). This condition included the clearance drawn as a graphic on the ND as well as the clearance presented as text to the right of the ND as illustrated in Figure 1. Both the current flight path (magenta) and clearance are displayed (green).

Graphics + Text + with updated SC214 UMs (G+T+updUM). This is identical to G+T except some UMs had been updated an were therefore updated for this condition.

Graphics + Text + ASD. This condition is the same as G+T conditions except an ASD was designed to provide additional altitude graphics, and was placed below the ND. Figure 2 illusrates the ASD.

Graphics +Integrated Text + ASD. The text was removed from the right of the ND and placed directly on the ND along with the graphics. The ASD was placed below the ND.

Dependent Variables

Pilot performance was measured with two dependent variables: response time to interpret the clearance and mean percent correct response. Each clearance was replicated four times. Two clearances were designed into scenarios to be correct and pilots were expected to ACCEPT the clearance based on instructions provided at the beginning of the testing. Two clearances were incorrect and pilots were expected to REJECT the clearance. Therefore the dependent variable for mean percent correct includes 1) mean percent correct accepts, and 2) mean percent correct rejects.

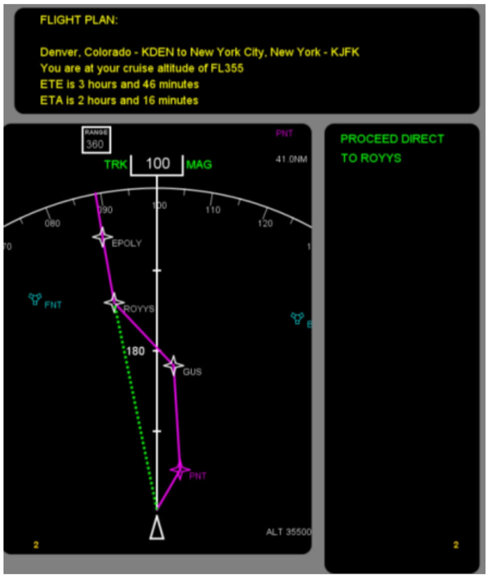


Figure 1. **G+T Format. Green Dotted Line Shows UM Graphic, Proceed Direct to ROYYS.**

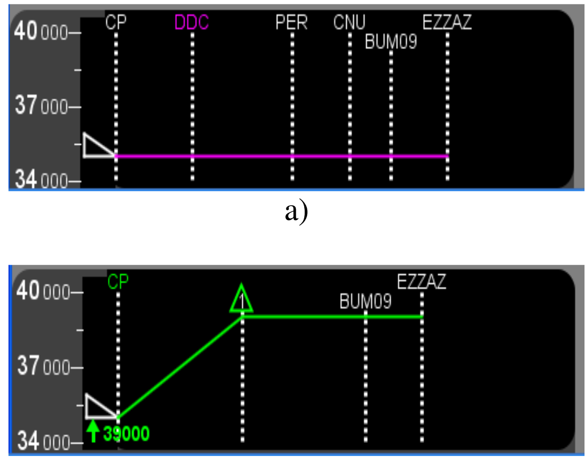


Figure 2. **Example of the ASD. Magenta line is current alitude path. Pilot toggles to clearance shown using green line.**

Subjects

Pilots were recruited from the Dayton region and and Cessna Aircraft Company, Wichita, KS. Pilots were screened for a minimum of 100 flight hours. A total of 66 pilots were tested across the various formats. Not all pilots saw all formats. Each pilot participated in two formats maximum. Pilot average age was 43.4 years. The average number of flight hours was 5,926, and 97.5% of the pilots held an instrument rating.

Hardware/Software

The experiment was controlled via a Hewlett Packard (HP) laptop (model Elitebook 8560p) and a portable 18 inch LCD monitor (model L1940T). Connected to the HP laptop was an external 19-key keypad (model FC K19U) data entry device.

The software was custom designed and developed using JAVA programming language and run in the NetBeans IDE. The software controlled the presentation of all display formats, timing, and data collection.

Procedure

The pilot sat in front of the screen with a small keyboard input device. The pilot pressed the 'Enter' key when ready to evaluate the current flight situation, and a timer was started. The pilot reviewed the flight plan and map until they understood the current situation. Once they became familiar with the situation, they pressed the Enter button again and the flight situation study time was recorded. Next, the clearance was displayed, the flight plan text was removed, and the response timer was started again. Pilots had the option of displaying the flight plan again by pressing the backspace (BS) button. After reviewing the clearance, the pilot either accepted or rejected the clearance by pressing 1 or 3 respectively on the external keypad. The judgment was based on the information provided to the pilot through the flight plan, navigation display, UM text and when relevant the ASD.

The pilot was instructed to accept the clearance if it directly matched the flight plan they had studied or if the clearance called for a deviation from the flight plan but led them to the same destination or future waypoint on the original plan. The pilot rejected the clearance if it did not match the flight plan or sent them on a path that did not lead to their destination. In addition the subject would reject the clearance due to excessive additional distance flown (even if directed back to destination), inappropriate altitude for phase of flight, and altitude mismatch. For example the clearance might have required flying to a waypoint already passed, or to a waypoint off the flight plan and in the wrong direction. If the pilot felt that a scenario and clearance were confusing they were asked to note the scenario number, but to move forward and respond to the clearance. Pilots were asked to respond as they would during actual flight by accurately evaluating the clearance in a timely fashion followed by rapidly indicating an intent to comply (accept) or their concern about the acceptability or validity of the clearance by responding in the negative (reject).

Results and Discussion

An ANOVA was conducted using the procedure PROC MIXED to deal with unequal N across conditions. The analysis was conducted with two IV variables: FORMAT (text and graphic conditions) and RESPONSE (Correct Accept or Correct Reject). The use of RESPONSE as an IV in the analysis provided the ability to analyze whether the pilots correctly accepted or rejected clearances based on FORMAT type.

The exploratory experiment was designed to evaluate performance differences between text and graphic conditions by comparing the baseline TEXT condition to the G+T condition. Figure 3 illustrates the trend of increased response time as the number of elements increase for both TEXT and G+T. The increase in MRT for TEXT, that is the slope, is much greater as the number of elements increase compared to that of G+T. The decrease of MRT for the nine element clearance for G+T is most likely due to the fact that pilots tended to reject these clearances more than accept them. (Correct rejections were 94.17% while correct accepts were 45.83%). Correctly rejecting a clearance was always faster than correctly accepting a clearance. The results indicate that a hybrid of graphics and text is not needed for one and two element clearances. However, for three elements and above graphics and text improved MRT.

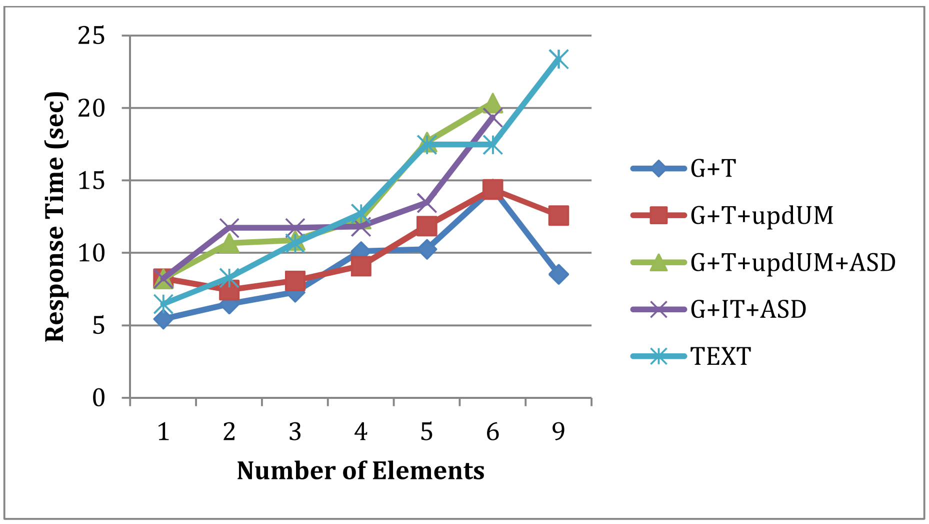


***Figure 3*. Mean response time as a function of format and number of elements in a clearance.**

There are some performance differences among the graphic formats. Specifically when the ASD was included MRT increased. Pilots did not think the ASD was necessary as altitude was available on the ND. Figure 3 illustrates that the two conditions with ASD are grouped together across number of elements, while the G+T hybrid conditions (G+T+ updUM) are grouped with lower MRTs. TEXT and G+T+ASD results are very similar.

Figure 4 illustrates the mean percent correct across the graphic conditions and number of elements. Although the four element condition showed no significant differences, the p value was only .01 below the 0.05 criterion. In this case the trend is for G+IT+ASD to be significantly different than G+T. The five element clearance also showed a significant difference for G+IT+ASD compared to G+T and G+T+ASD. For six element and nine element clearances there was no difference across graphic formats. There is a need for additional research to ensure the difference in mean percent correct across graphic formats. There was a strong trend suggesting the need for additional research related integrating text (G+IT+ASD) onto the ND rather than separating text and eliminating the ASD.

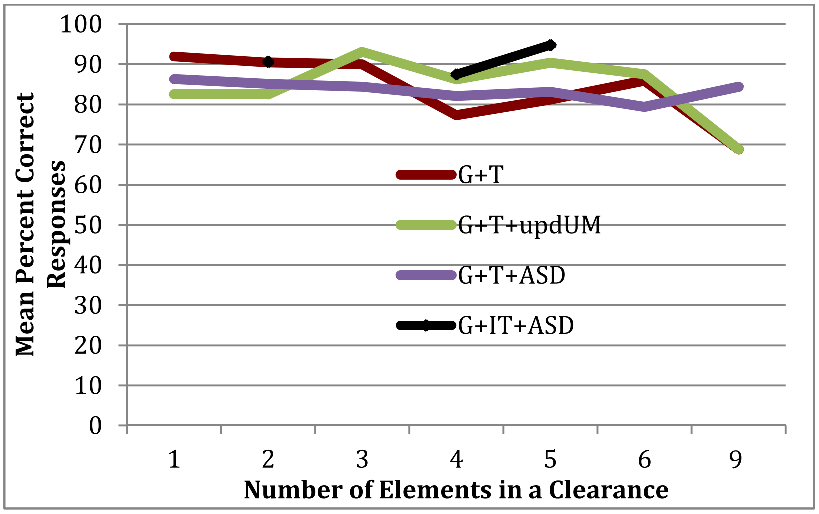


Figure 4. **Mean Percent Correct Responses to Accept a Clearance as a Function of Format.**

In summary results indicated that overall as the number of elements in a clearance increased the time required to interpret the clearance increased and errors increased. Hybrid conditions with graphics + text improved performance when there were 3 or more elements in a clearance. Graphics provided an opportunity for pilots to compare their mental model of their current and future aircraft positions to the graphic. Text alone required pilots to mentally project their current and future condition on to the ND.

Recommendation Examples

The primary result of this exploratory research was the development human factors recommendations. Examples of recommendations are presented in Table 1. Readers are referred to the report for more in-depth descriptions, rationale, and graphic examples.

Table 1
Example recommendations based on human-in-the-loop research.

Category	Recommendation
Coordinating Text with Graphics	When text and graphics are presented separately, there should be symbols or other design methods that illustrate the one-to-one match of the text and the coordinating graphic.
Distinguish between simultaneous versus sequential UMs and DMS.	The graphic and text should clearly indicate simultaneous versus sequential operations.
Rejoin Route Graphics	If the intended meaning of this UM is to allow pilots to rejoin the route at their discretion as long as it is before the POSITION, then a single green horizontal line at POSITION is effective at providing the limit by which they must rejoin. However, this would allow the pilot to rejoin as soon as 30 seconds and still be in compliance. If the intention is more specific as to when to begin the rejoin after an offset, than a green horizontal line with a shaded region indicating the zone in which they may rejoin reduces ambiguity.
Current Setting of Range Level for DataComm Graphic Displays	When a new clearance appears the range of the ND may not be at the correct setting to view the graphic appropriately. Changing the range automatically may confuse the user. The range of the DataComm display should remain at the last setting the user applied.
UM visibility after pilot decision.	The graphic and textual UMs should remain visible after a pilot decision (WILCO/Unable) until the pilot makes an action to clear the clearance. Once the decision is made there should be a visible indication of the decision selected. Examples include removing or graying the WILCO/Unable selection or making a change in how the text or graphic UM is displayed.

Conclusions

This research evaluated text and hybrid text and graphic concepts to develop human factors (HF) recommendations for specialists who participate in certification of new and modified flight deck DataComm displays, and as a potential update to AC 20-140, Guidelines for Design Approval of Aircraft Data Link Communication Systems Supporting Air Traffic Services (ATS).

This is one of the first studies to directly compare text display of clearances to graphic and/or hybrid presentations of graphics and text to the flight crew. The research findings indicate that when three or more elements are specified in a clearance, flight deck presentation methods that include graphics and text result in better human performance outcomes than text presentation alone.

Acknowledgements

The FAA NextGen Human Factors Division (ANG-C1), coordinated the research requirement and its principal representative acquired, funded, and technically managed execution of the research services described in this report.

In addition to the research team, Cessna Aircraft Company provided direct support for the completion of this project. Cessna provided access to highly trained pilots as volunteers to serve as research subjects. We also thank them for their support in providing feedback during data collection and during review of the human factors recommendations.

References

Cardosi, K., Lennertz, T., Donohoe, C. (2010). *Human Factors Research Plan for Flight Deck Data Communications*. Human Factors Research and Engineering Group. Boston MA: USDOT Volpe National Transportation Systems Center.

Gallimore, J.J., Shingledecker, C., Tsang, P.S., Oh, C., and Kiss, Stephen B., (2011). Interim Literature Review Report for Flight Deck Display and Control Requirements Human Factors Study. Technical Report prepared for FAA, NextGen Advanced Concepts and Technology Development, Human Factors Division (ANG-C1), DTFAWA-10-A-80021., Wright State University, Dayton, OH 45305

Gallimore, J.J. Kiss, S.B., Munoz, R.D., Oh, C., Crory, T., Ward, B., Green, R., Shingledecker, C., and Tsang, P. (2013). DataComm – Display Alternatives for the Flight Deck: Overview and Human Factors Recommendations. Final Tech Report Vol 1 & 2. FAA NextGen Advanced Concepts and Technology Development, Human Factors Division (ANG-C1), DTFAWA-10-A-80021. Wright State University, Dayton OH, 45435. *Available at www.hf.faa.gov.*

Hahn, E. C., & Hansman, R. J. (1992). Experimental Studies on the Effect of Automation on Pilot Situational Awareness in the Datalink ATC Environment,(SAE Tech. Report No. 922002). Warrendale, PA: Society of Automotive Engineers International

Prinzo, O. V. (2003). Pilot's Visual Acquisition of Traffic: Operational Communication from an In-Flight Evaluation of a Cockpit Display of Traffic Information. *The International Journal of Aviation Psychology*, 13(3), 211-231.

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional Models of Multitask Pilot Performance Using Advanced Display Technology. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 45*(3), 360-380.

CONTROLLER – PILOT COMMUNICATIONS IN THE PRESENCE OF ASYNCHRONOUS UAS RADAR SURVEILLANCE DATA

Xiaochen Yuan
University of Waterloo
Waterloo, ON, Canada

Jonathan Histon
University of Waterloo
Waterloo, ON, Canada

Catherine Burns
University of Waterloo
Waterloo, ON, Canada

Steven Waslander
University of Waterloo
Waterloo, ON, Canada

Reza Dizaji
Raytheon Canada
Waterloo, ON, Canada

Integrating Unmanned Aircraft Systems (UAS) into controlled airspace will create challenges for the pilots and controllers who need information about the UAS. This paper presents a preliminary study of the effect of differential time delays, or asynchrony, in the distribution of UAS surveillance information to controllers and pilots. Effects on controller-pilot communication were observed through 6 distinct measures of both objective performance and subjective self-evaluation. Larger time delays had an observable impact on all of the observed measures; comparison of pilot and controller results showed that the operator with the most updated information consistently experiences less frustration and feels the communication were more effective.

Driven by the huge profit opportunities in the Unmanned Aircraft Systems (UAS) market, the use of UAS is gradually shifting from exclusively military applications to civil applications. The expanded use of UAS will require integration of their operations into airspace actively managed by an air traffic controller ("controlled airspace"). Recent news stories highlight the concern generated by the integration, planned or ad hoc, of UAS into controlled airspace (Hurtado, 2013). Incidents reported to NASA's Aviation Safety Reporting System highlight concerns about the different knowledge/information of nearby UAS to pilots and/or controllers (NASA, 2013). In addition, new surveillance capabilities provide opportunities to collect and distribute data on smaller and/or non-cooperating UAS to both pilots and controllers.

The ability to more widely distribute radar surveillance data on UAS and other objects to both controllers and pilots raises a number of Human Factors challenges. Yuan et al. (2012) identified three important challenges. Differences in how the information is distributed to pilots and controllers can create differential time delays; this can result in pilots and controllers collaborating while viewing information that is a different "age" despite coming from a common source. The potential difference in the "time age" of shared information has been labeled as *Information Asynchrony* (Yuan et al., 2012). While there has been significant previous work on controller-pilot collaboration, there does not appear to be much data available on how differential time delays in a common information source affect the communication and collaboration between pilots and controllers.

This paper presents a preliminary study of the effect of asynchronous information on pilots and air traffic controllers' communication. The paper first reviews related work on the impact of time-delays on controller-pilot communications, briefly describes the design of the experiment, and presents the results of the study.

Background

Ambiguity, errors, and miscommunications between pilots and controllers are all potential causes of accidents (Morrow, Lee, & Rodvold, 1993). An important factor affecting communications is the presence of time delays (Morrow et al., 1993). Rantanen, McCarley and Xu (2004) have investigated the effect of systemic audio delay (AD) and variable pilot delay (PD) on controller performance and workload. Asynchronous information already creates challenges around communicating location of convective weather and confusion generated by the presence of time delays has prompted the National Transportation Safety Board to issue safety alerts about the use of NEXRAD mosaic image by pilots (NTSB, 2012). Day et al. (1999)studied the effects of delayed visual feedback and found that it produced oscillations in control movements and targeting exercises. Outside of the air traffic control domain, Kraut, Gergle and Fussell (2002) have examined the effect of time delays in a contrived jigsaw puzzle collaboration task with a shared visual display. Introducing a delay of as little as 3 seconds in the task was reported to impact performance and "in many cases rendered the shared visual space useless" (Gergle, Kraut, & Fussell, 2006).

Controller-pilot communications are important for maintaining consistent and accurate mental models of the traffic situation for both the pilot and the controller (Mogford, 1997). The shared mental model between pilots and controllers will include weather, traffic, intent and affective states (Farley & Hansman, 1999). Farley and Hansman (1999) experimentally studied the effects of increased sharing of weather and traffic data between pilots and controllers; however, this previous work assumed that information sharing would be instantaneous and did not examine the effects of differing time delays in access to shared information.

New technologies, such as System Wide Information Management (SWIM) (Meserole & Moore, 2007), are creating the opportunity to more broadly share information between pilots and controllers (Ulfbratt & McConville, 2008). Yuan et al. (2012) have developed a model of future operations incorporating SWIM; the model was used to identify potential Human Factors challenges resulting from the expanded distribution of surveillance data on non-cooperative objects. A key challenge identified was the potential of pilots and controllers coordinating resolution actions while dealing with asynchronous information. This is similar to challenges in current operations with communicating about convective weather (Brown, 2007).

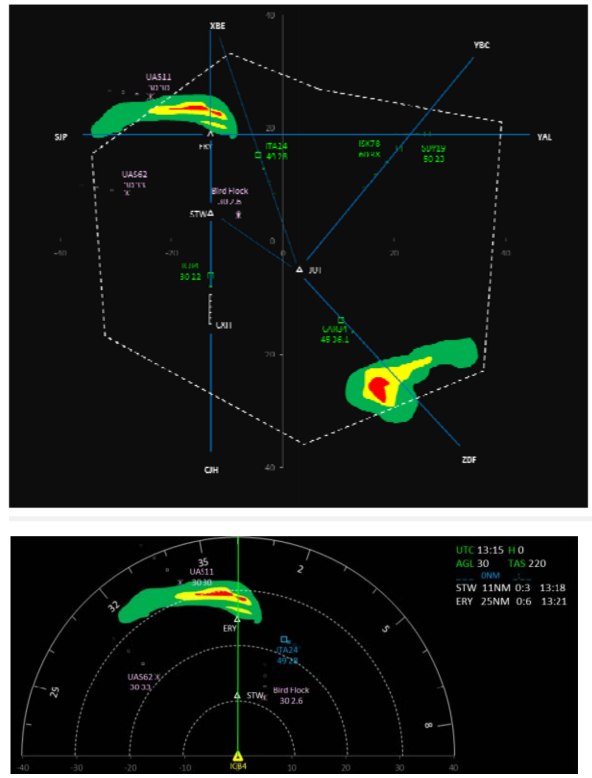


Figure 1.
(Top) Controller View, (Bottom) Pilot View

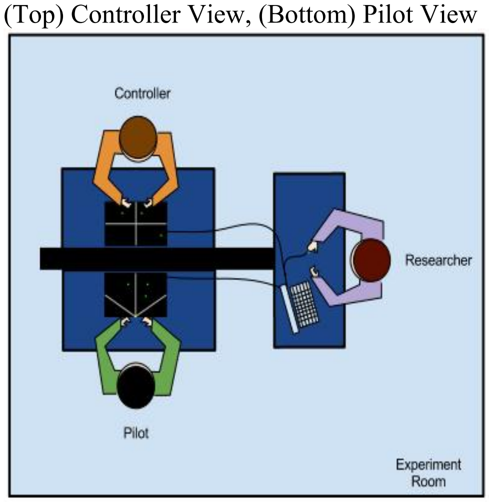


Figure 2.
Experiment Room Setting.

Experiment Probing the Effect of Asynchronous Information on Controller-Pilot Communication

A simple experiment was designed as an initial examination of the effect of asynchronous information on controller-pilot communication. In the experiment, participants were shown static pictures of a radar surveillance display (controller participant) and a primary navigation display (pilot participant) (Figure 1). Their tasks were to observe the displays, identify potential conflicts and communicate with each other to resolve the conflicts. Relevant data, including the communication time and subjective mental status, were collected. The experiment manipulated the amount of time delay between the information presented to the controller participant and the pilot participant between 0 and 10 minutes.

Experimental Setup, Tasks and Scenario Design

The task was setup to resemble current controller-pilot voice communications. A divider was placed between the two participants (Figure2) allowing them to communicate verbally but without being able to see the other person or any gestures.

In each trial, participants were asked to evaluate their radar surveillance display/navigation display and communicate with each other in order to resolve any conflicts presented. . Controllers were notified of the general traffic situation in their controlled sector by providing a written briefing prior to the trial. Each group of participants performed five experimental trials; each trial used a scenario randomly selected from a pool of previously generated scenarios. A delay interval of (0, 0.5 1, 5, 10) minutes was applied to the information for either the controller or pilot display. To minimize potential learning effects, the sequence of time delays in the experimental trials was random.

Ten different scenarios were designed; these formed a pool of five takeoff scenarios and five landing scenarios. This was done to minimize any effects of the details of the traffic situation on the results, while still presenting participants with novel and engaging situations in each trial. For each scenario, the common elements shown to both participants included traditional aircraft, as well as depictions of non-cooperative objects such as birds, weather and UAS. One or more impending conflicts between the pilot's aircraft and a UAS/weather condition/birds were embedded in each scenario. For each trial, once a scenario was selected from the pool, the required time delay was applied to the non-cooperative objects in the selected scenario.

Participants and training

Access to trained professionals was not possible; instead, students from a local university were recruited as participants. Consequently, the participants were expected to be not as familiar with air traffic control and piloting operations. However, for the purpose of the intended task, it was felt that the core goal of communicating to negotiate a coordinated resolution to a conflict did not require specialized knowledge. Since the participants were naïve, multiple training trials were provided to make sure that the participants were capable of reading the display, analyzing potential conflicts and communicating with each other. In the display training, participants were taught to recognize the legends (aircraft icon, non-cooperative objects icon, etc.) and understand the information presented on the displays. Training on how to communicate was provided; 4 static displays were used to illustrate to participants how to communicate regarding potential conflicts.

Participants were comprised of 12 groups (12 female and 12 male participants) with an average age of 29.5 years old. Participants were assigned to the pilot and controller roles randomly at the start of the experiment. And participants did not previously know each other. In addition, in order to eliminate the effect of gender, there were three "male – male" groups, three "female – female" groups and six "female – male" groups.

Data Collection

Both objective and self-reported data was collected. At the end of each experimental trial, the participants completed a post-trial questionnaire with four 10-point Likert scales asking participants to rate their self-assessed *performance, communication effectiveness, frustration level* and *trial difficulty.* As well, the experimental trials were audio recorded for the purpose of analyzing communication time and clarification information.

Results

Objective Data

Communication time. Communication time refers to the time from the beginning of the trial until the participants reached an agreement on a resolution action. Figure 3 (a) presents the average communication time observed for each time delay interval. Error bars in the figure represent the standard error. The general trend indicates that the communication time increases when someone has more updated information (either the pilot or the controller). A repeated measures ANOVA analysis was also performed. F (4, 36) = 2.54, p = .057 > .05 which indicates that there is no significant relationship between communication time and the delay interval.

Number of clarification times. A clarification was defined as the moment that 1) one party of the pair asks the other party to either confirm and/or describe one or more particular object(s), and/or 2) one party of the pair disagrees with the other on the description/position/information of one/more particular object(s). Therefore, if one party merely does not hear the description clearly and asks for more explanation, it was not counted as a clarification. Figure 3 (Right) demonstrates the relationship between the average number of clarifications and the delay intervals. The overall trend indicates that an increase of the time delay increases the number of the clarification times. Repeated measure analysis shows that the repeated scenarios have a significant impact on the number of clarification times (F (1, 9) = 4.080, p = .008 < .05).

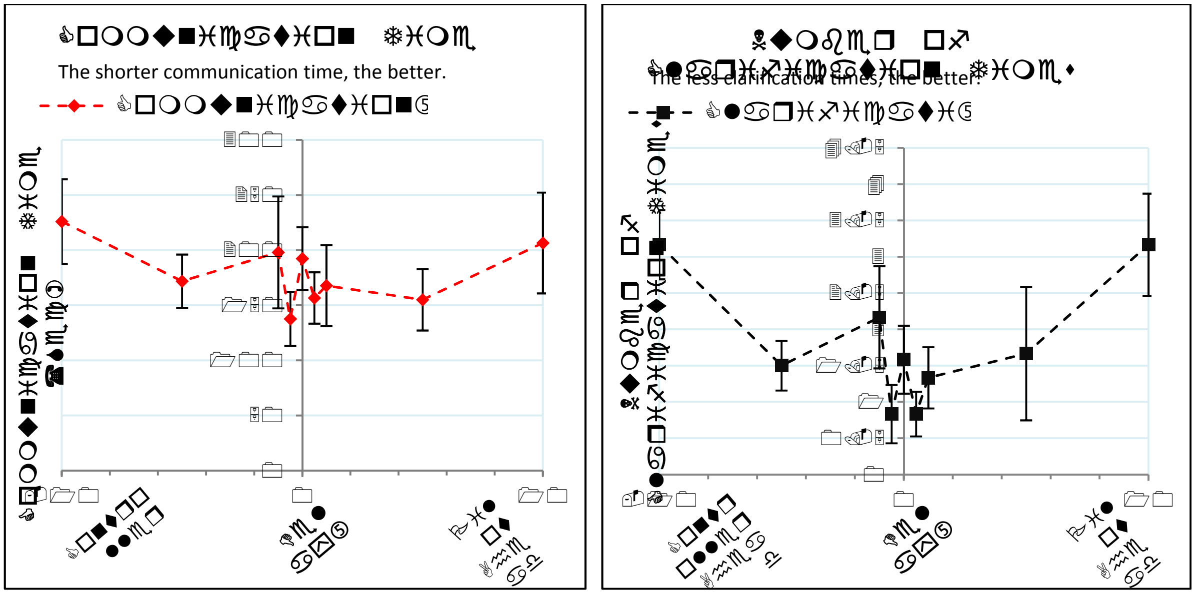


Figure 3.
Mean value plots of the Clarification Times (Left) and Communication Time (Right).

Subjective Data

The main resource of subjective data is the self-rate measurements collected from the post-trial questionnaires. The 4 measures were *performance*, *communication effectiveness*, *frustration*, and *trial level of difficulty*. Checks were made for learning effects and the data showed that the ratings were generally consistent, independent of trial number.

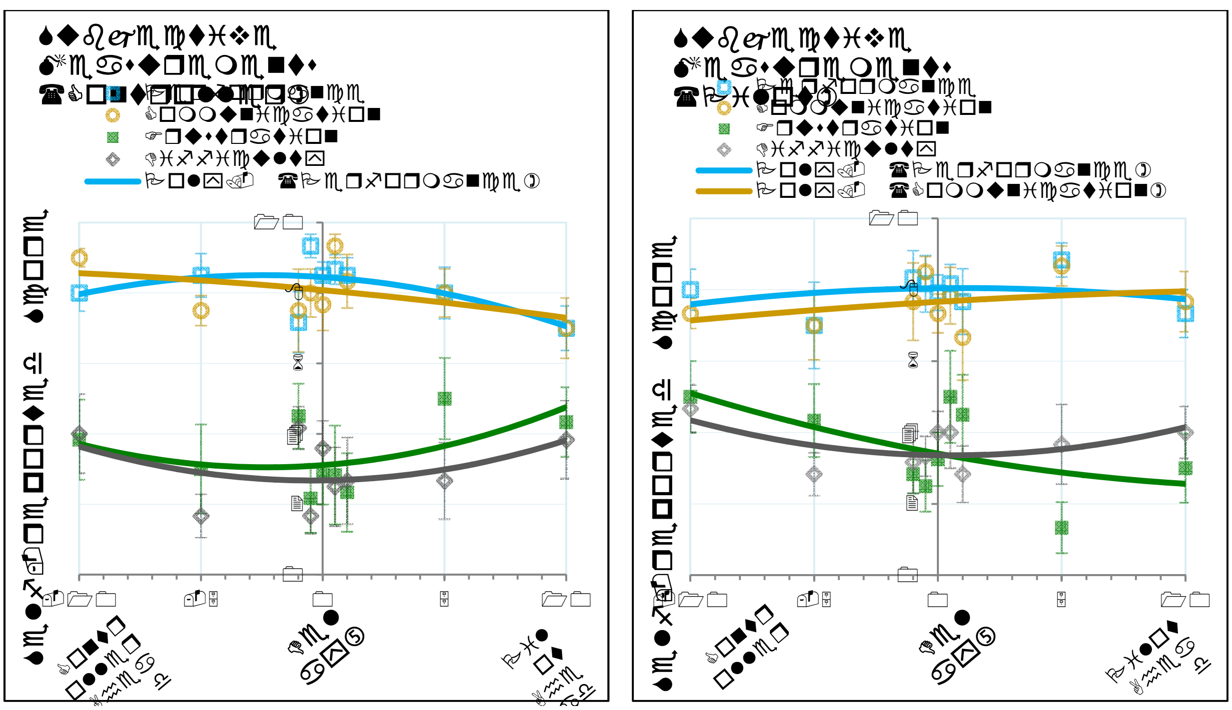


Figure 4.
Mean value plots of subjective measurements for controller (Left) and pilot (Right).

Figure 4 shows two mean value plots of the 4 measures for the controllers and pilots respectively. In general, when there is no delay, both pilots' and controllers' self-reported performance scores (blue line) are the highest and they feel the least difficulty (grey line) of the trial. Figure 4 also shows that when the pilots have the most up-to-date information, the controllers feel more frustration (green line), and the communication is less effective (yellow line). And vice versa.

In order to better understand this result, a further analysis were performed to compare the 4 measures on the same scale of “who is ahead” in relation with the score/level of the measurements (Figure 5). It indicates that the party, who has the most updated information, thinks the communication is the most effective and the least frustrating.

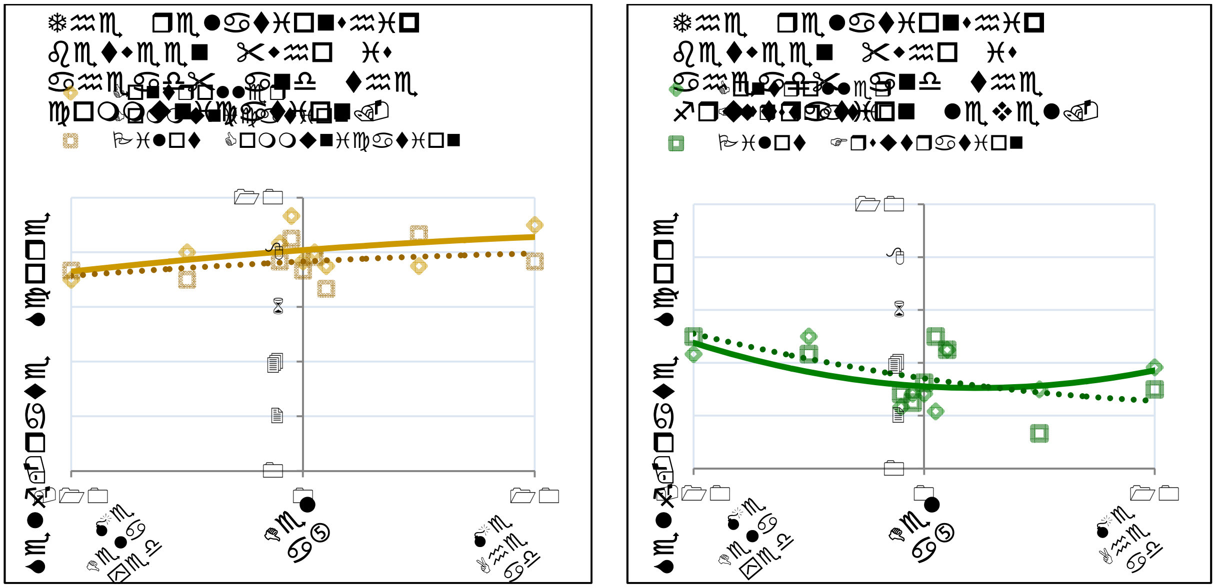


Figure 5.
The relationship between “who is ahead” and communication (Left), “who is ahead” and frustration (Right).

Discussions, Implications & Future Work

There are several key implications from the results presented above. Delay time had a clear effect on the communication performance. Long delays on the order of 10 minutes produced increased frustration and need for clarifications, and longer conflict resolution times. While 10 minutes is an unrealistic delay time for distributing object locations, it is on the same order of the time delays that are experienced when communicating about weather in current operations. Future work could consider differential update rates as a source of the differential time delays. In this study, asynchronous information on UAS, birds, and weather were all updated at the same rate. In operations, however, the dynamics of these objects are not exactly the same and there may be advantages to having different update rates. For slow-moving objects, such as weather and broad areas of bird activity, rapid updates may be perceived as unnecessary as reducing update rate is one method of reducing costs and bandwidth requirements.

In addition, it was obvious that the operator receiving the most up-to-date information had a better communication experience. Results showed they felt less frustration and were more likely to feel the communication was effective. This result is different from our original hypothesis, which was that with an increase in the time delay, both parties would uniformly feel more frustrated and less effective at communication.

There are several limitations to the study that restrict the implications that can be drawn, particularly for shorter delay times; however, these limitations can be addressed in future studies. Effects of differential time delays at the shorter durations were not as pronounced due to the noise in the data. It is thought that the use of static pictures eliminated important time pressure factors and did not present participants with a dynamic vision of the situation. The lack of motion could have affected the participants’ ability to make precise predictions and decisions and masked differences in the effects of shorter delay times.

Future research will narrow the delay window to a smaller range that is more likely to represent the time delays that would be observed for transmitting object position data. The study will be repeated in a part-task dynamic simulation environment and it is hoped that professional controllers and pilots can be recruited.

Acknowledgements

The researchers express appreciation for the support and funding of this work jointly by the Canadian National Science and Engineering Research Council Collaborative Research and Development Grant (#CRDPJ-398768-2010) and Raytheon Canada Limited. The material presented in this paper represents the views of the authors and do not necessarily reflect the views of the funding agencies/company.

References

Brown, D. (2007). Say Again #71: Weather Radar, Published January 31, 2007. Retrieved March 1, 2013 from http://www.avweb.com/news/sayagain/194130-1.html

Day, P., Holt, P., & Russell, G. (1999). Modeling the effects of delayed visual feedback in real-time operator control loops: a cognitive perspective. *Proceedings of the XVIII European annual conference on human decision making and maual control* (pp. 70–79). Loughborough, UK: Group D Pblications Ltd.

Farley, T., & Hansman, R. (1999). An experimental study of the effect of shared information on pilot/controller re-route negotiation. *International Center for Air Transportation*, (January).

Farley, T., Hansman, R., Endsley, M. R., Amonlirdviman, K., & Vigeant-Langlois, L. (1998). *The Effects of Shared Information on Pilot-controller Situation Awareness and Re-route Negotiation*. Massachusetts Institute of Technology, International Center for Air Transportation.

Gergle, D., Kraut, R., & Fussell, S. (2006). The impact of delayed visual feedback on collaborative performance. *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1303–1312).

Kraut, R., Gergle, D., & Fussell, S. (2002). The use of visual information in shared visual spaces: Informing the development of virtual co-presence. *Proceedings of the 2002 ACM conference on Computer supported cooperative work* (pp. 31–40).

Meserole, J., & Moore, J. (2007). What is System Wide Information Management (SWIM)? *Aerospace and Electronic Systems Magazine, IEEE*, *22*(5), 13–19.

Mogford, R. H. (1997). Mental models and situation awareness in air traffic control. *The International Journal of Aviation Psychology*, *7*(4), 331–341.

Morrow, D., Lee, A., & Rodvold, M. (1993). Analysis of problems in routine controller-pilot communication. *The International Journal of Aviation Psychology*, *3*(4), 285–302.

NTSB. (2012). *In-Cockpit NEXRAD Mosaic Imagery* (pp. 1–3). Retrieved from http://www.ntsb.gov/doclib/safetyalerts/SA_017.pdf

Rantanen, E., McCarley, J., & Xu, X. (2004). Time delays in air traffic control communication loop: effect on controller performance and workload. *The International Journal of Aviation Psychology*, *14*(4), 369–394.

Ulfbratt, E., & McConville, J. (2008). *Comparison of the SESAR and NextGen-Concepts of Operations. NCOIC Aviation IPT.*

Yuan, X., Histon, J., Waslander, S., Dizaji, R., & Schneider, C. (2012). Distributing non-cooperative surveillance data: A preliminary model and evaluation of potential use cases. *2012 Integrated Communications, Navigation and Surveillance Conference* (pp. F5–1–F5–10). IEEE. doi:10.1109/ICNSurv.2012.6218397

LOOKING FOR CHUCK AND PT: AN EVIDENCE-BASED APPROACH TO ASSESSING HELICOPTER PILOTS

Paul Dickens
Core Business Psychology Ltd.
Edinburgh, Scotland UK

This paper describes the use of an evidence-based approach to the assessment of commercial rotary-wing pilots. Following the four category protocol set out by Briner and Rousseau (2011), a robust psychological assessment process was developed covering intellectual ability, mental wellbeing, the Big 5 personality dimensions and critical incident analysis. Findings from each component are discussed, particularly the findings on the personality characteristics using the Big 5 dimensions. Findings of high Conscientiousness and very low Neuroticism were shown to mirror similar findings in the literature, while a higher than expected incidence of social withdrawal needs more investigation. Implications for future assessment processes and research are described.

Having received an assignment from a North Sea helicopter operator for in-depth psychological assessment of helicopter pilots and co-pilots they were recruiting, we decided to take an evidence-based approach following Briner and Rousseau's (2011) suggestions. This included their suggested 4 key information sources;

- Practitioner expertise and judgment
- Critical evaluation of the best available research
- Evidence from the local context
- Perspectives of those who may be affected by the intervention decision

Activities under each of these headings are described. Based on each of these sources, and in alignment with the company's process used elsewhere in the world, we settled on an assessment process that met both our desire to be evidence-based and the client's requirements for a robust, cost and time effective process.

During our review of the relevant literature and best available research, we found few studies that defined either a robust assessment process for selecting rotary-wing pilots or a clear indication of the desirable psychological characteristics. Two papers (Grice (2006) and Grice and Katz (2008)) came closest to providing a clear template of the personality dimensions typical of pilots in a military (US Army Air Corps) setting, and they differentiated the personality characteristics of pilots using the Big 5 dimensions across a number of rotary-wing mission platforms including attack, transport and utility aircraft.

The assessment process

The process combined the following elements:

- A measure of intelligence – Raven's Advanced Progressive Matrices (Raven, Raven and Court (1998))
- A measure of mental wellbeing – the GHQ28 (Goldberg and Williams (1988))
- A measure of personality – the Big Five Inventory (John, Donahue and Kentle (1991))
- An in-depth structured interview including critical incident analysis (Flanagan (1954)) and a strengths-based inquiry (Linley et.al. (2007)). The interview questions were developed using an evidence-based approach based on research into the characteristics of safe and unsafe pilots (FAA (2008)).

Sample characteristics

Table 1 below shows the number and gender of pilots in the study sample at 1st January 2013. By the time of presentation this sample will be larger. Comparison is made to the most recent statistical account available of the number of pilots holding the basic qualification required by the company – Commercial Pilot's License (Helicopter) (CPL (H)) – as listed by the UK Civil Aviation Authority (CAA 2008).

Table 1.
Totals and Gender Balance of the Sample

	Sample	%	2008 CAA	%
Female	3	4%	23	4%
Male	53	96%	617	96%
Total	56	100%	640	100%

Table 2. shows the average age and flying hours of the sample, again with a comparison to the 2008 CAA summary statistics.

Table 2.

Average Age and Flying Hours of the Sample.

	Average	Range	2008 CAA	Range
Age	36	21-55	35	20 - 61
Hours	2250	100 - 7500		

Findings from the assessment process

Raven's Advanced Progressive Matrices

This test is recognized as one of the best assessment tools for assessing general intellectual capability. It is non-verbal and free from educational, cultural and gender bias. The version used here was developed to differentiate between people of superior intellectual capability. The test comprises two parts, Set 1 which consists of 12 items and is given untimed, acting as a training set. Set 2 consists of 36 similar items of increasing difficulty, and has a time limit of 40 minutes.

The average score for this sample on Set 1 was 11.08, SD 1.09, and on Set 2 the average score was 24.92, SD 4.71. On Set 1the range of scores was 8 to 12, and on Set 2 12 to 34. Most pilots are therefore of average intelligence or above, and low scores on Set 2 (i.e. less than 16) were an exclusion criterion for the selection process.

GHQ 28

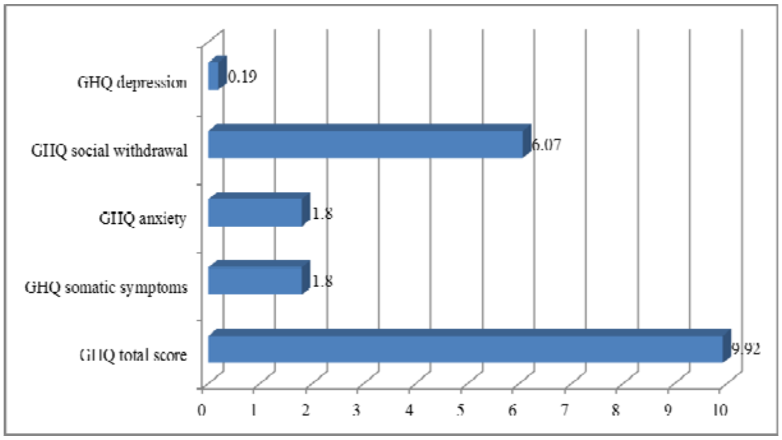
This is a well established screening tool to detect those likely to have, or to be at risk of, developing psychiatric disorders. Based on the full GHQ, the version used here measures 4 factors;

- Somatic symptoms
- Anxiety
- Social dysfunction
- Severe depression and suicidal tendencies

The GHQ 28 is scored two ways on the four point scale of symptom frequency – the so-called GHQ scoring (0011) and a Likert type scale (0123). "Caseness"- that is the likely presence of significant psychiatric morbidity - is a score of 4 or more using GHQ scoring, or 23/24 on the Likert scale. Figure 1 below shows the average scores on the total scale, and on each of the four component scales.

Figure 1. shows that pilots in general do not reach "Caseness", and comparison to the scores of comparable general populations shows a low incidence of psychiatric symptoms. In particular the scores on Severe Depression and Suicidal Tendencies are – thankfully – very low. By contrast the scores on Social Dysfunction, which includes social and work performance, are slightly higher than expected and significantly different from the other subscales scores. This may be attributed to the fact that almost all of the pilots sampled were actively looking for employment – hence the assessment – and often were unemployed or coming to the end of a previously very active military career.

Figure 1.
Average total and subscale scores on the GHQ.



Big Five Inventory

Because of the robustness of the research findings concerning the usefulness of the Big 5 model of personality, we included a direct measure of the 5 dimensions as part of the assessment process. We chose to use the Big Five Inventory (BFI) (John, Donahue and Kentle (1991); John, Naumann and Soto (2008). The reasons for this were that the BFI possesses robust statistical properties and is quick to complete and score, allowing questioning of answers during the interview process. Figure 2 below shows the average scores for this sample on each of the 5 dimensions, with a comparison expected score for the general population.

The finding in this sample mirror those obtained by Grice (2006). Grice used the NEO-PI to measure the Big 5 dimensions so a direct comparison is not possible, although Table 3. below shows the direction of the scores on each of the Big 5 classified into Low, Average or High. There are consistencies in these findings. Typically both samples of rotary-wing pilots show a lower than average score on Neuroticism – they tend to be calm, resilient and emotionally stable people who cope well under pressure. The other key characteristic is a very high level of Conscientiousness – they tend to be organized, methodical, procedure-driven and hard-working. The current sample also scores higher than expected on Agreeableness – they tend to be easy to get on with, well-mannered and committed to working well in a team.

Discussion

This paper is work in progress and the findings outlined here are preliminary data that may change as the sample grows. Similarly there is more information that can be gained from the assessment process through analysis of the critical incident reports and more particularly the strengths-based enquiry. Preliminary analysis of each individual's stated strengths shows a potential strong correlation with the Big 5 personality dimensions.

Figure 2.
BFI average scores of each dimension for the sample with comparison expected average scores for the general population.

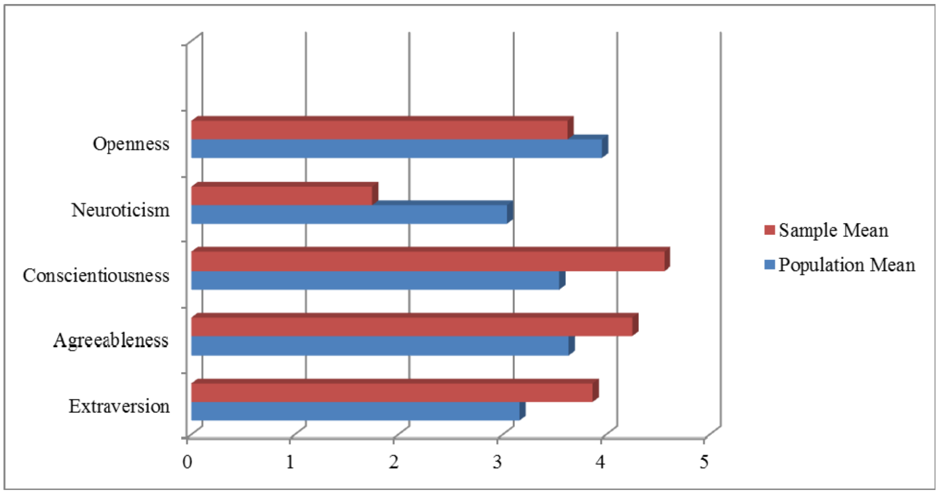


Table 3.
High level categorization of Big 5 personality dimensions compared to those put found by Grice (2006) in a US Army Air Corps sample

Factor	*US Army Total*	*US Army C&U*	*Current Sample*
Extraversion	*Average*	*Average*	*Average*
Agreeableness	*Average*	*Average*	*High*
Conscientiousness	*Average*	*High*	*High*
Neuroticism	*Low*	*Low*	*Low*
Openness	*Low*	*Low*	*Average*

Despite the initial status of the current findings it does appear that there are strong defining characteristics of rotary-wing pilots that are useful as a reference point in selection and training.

References (Level 1 Heading)

Briner, R.B. and Rousseau, D.M. (2011): Evidence-based I-O Psychology: Not There Yet. *Industrial and Organizational Psychology 4*, 3-22.

Civil Aviation Authority (2008). UK CAA Flight crew licence age profile as at 1 January 2008. London, England: Civil Aviation Authority. Retrieved from http://www.caa.co.uk/docs/175/UK CAA Flight Crew Licnece Age Profile as at 1 January 2008v2.pdf.

Federal Aviation Administration (2008). *Pilot's Handbook of Aeronautical Knowledge.* Oklahoma City, OK: US Dept. of Transportation Federal Aviation Administration, Airman Testing Standards Branch.

Flanagan, J.C. (1954): The critical incident technique. *Psychological Bulletin 51,* 4, 327-358.

Ganesh A. and Joseph C. (2005): Personality studies in aircrew: An overview. *Indian Journal of Aerospace Medicine, 49*, 1, 54-62.

Goldberg, D. and Williams, P. (1988) *A Users Guide to the General Health Questionnaire.* Windsor, UK: NFER-Nelson Publishing Company Ltd.

Grice, R.L. (2006): *Personality profiles of experienced US Army aviators across mission platforms*. Ph.D. Thesis. Lynchburg VA: Liberty University.

Grice, R.L. and Katz, L.C. (2008) Personality Profiles of U.S. Army Initial Entry Rotary Wing Students Versus Career Aviators: *USARI Technical Report No. 1208*. Arlington VA: United Sates Army Research Institute.

John, O.P., Donahue, E.M., and Kentle, R.L. (1991). *The Big Five Inventory – Versions 4a and 54.* Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, O.P., Naumann, L.P., and Soto, C.J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O.P. John, R.W. Robins & L.A. Pervin (Eds.) *Handbook of personality: Theory and research* (pp. 114-158). New York, NY: Guilford Press.

Linley, P.A., Maltby, J., Wood, A.M., Joseph, S., Harrington, S., Peterson, C., Park., N., and Seligman, M.E.P. (2007) Character strengths in the United Kingdom: The VIA Inventory of Strengths. *Personality and Individual Differences 43*, 341-351.

Raven, J., Raven, J.C., and Court, J.H. (1998) *Manual for Raven's Progressive Matrices and Vocabulary Scales.*San Antonio, TX: Pearson.

THE PSYCHOMOTOR VIGILANCE TEST: SOURCES OF STATE AND TRAIT VARIANCE

Hans-Juergen Hoermann, PhD, FRAeS
German Aerospace Center (DLR)
Institute of Aerospace Medicine
Department of Aviation and Space Psychology, Hamburg/Germany

Within the context of pilot and air traffic controller selection tests the Psychomotor Vigilance Test (PVT, Dinges & Powell, 1985) was evaluated for its underlying sources of variance. The PVT is a simple visual reaction time task, which is well established as a measure of alertness during sustained operations. It provides scores for mean reaction times and number of lapses. While the PVT has proven sensitivity for temporary states of fatigue and other stressors in within-subjects designs, validation studies are lacking to examine it for potential sources of trait variance, which could lead to confounding effects. This paper presents results from a validation study of the PVT with N = 247 air-traffic controller applicants. The PVT was administered in the morning before and in the evening after the selection tests. Lapses and mean reaction times show a different pattern of stability coefficients and inter-correlations with other tests. PVT-lapses appear to be more sensitive to state changes than mean reaction times. However, the change scores of the lapses seem to be confounded with individual personality traits. The PVT reaction time scores show significant correlations to selection tests of psychomotor skills, response orientation and vigilance. Implications for using the PVT as a potential selection test for pilots or air-traffic controllers are discussed.

The Psychomotor Vigilance Test (PVT, Dinges & Powell, 1985) has been widely used in sleep deprivation experiments to assess the impact of fatigue on performance (Elmenhorst et al., 2009). The PVT is a visual reaction time task asking subjects to respond as soon as possible to simple stimuli usually presented on a computer screen. The standard test length is 10 minutes with about 60 to 70 signals.

The test has proven empirical validity and sensitivity as a measure of effects of fatigue and other stressors (e.g. Loh et al., 2004; Elmenhorst et al., 2009; Basner & Dinges, 2011). Furthermore, it was stated that the PVT is “ideal” for repeated use in within-subjects designs because of its short testing time and independence of learning effects and aptitude differences, which often affect other cognitive tests (Dorrian, Rogers & Dinges, 2005; Dinges et al., 2012). However, studies are lacking, which examine the PVT for potential sources of trait variance that could result in confounding effects.

The goal of this study is to examine the temporal stability of different PVT scores and to examine the PVT for potential underlying sources of trait variance in the context of standard aptitude tests for air-traffic controller selection. Additionally, interindividual differences with respect to the subjects’ vulnerability to effects of test fatigue and time of day are assessed.

Methods

Subjects

N = 247 applicants (88 or 35% female) for the air-traffic controller (ATCo) ab-initio training of the DFS in Germany participated in the study. For 207 subjects complete data sets were available with two points of measurement for the PVT and the Subjective Fatigue Checkcard (Samn & Perelli, 1982). The additional 40 subjects had only one point of measurement. The subject's age range was 18 to 25 with an average of 19.8 years.

Measurements

Psychomotor Vigilance Test (PVT). A ten-minute version of the PVT was administered. Different performance scores were calculated:

- Overall mean reaction times ($PVT_{(1/2)}$-MRT)
- Mean reaction times for the 10% fastest responses ($PVT_{1/2}$-RTF)
- Mean reaction times for the 10% slowest responses ($PVT_{1/2}$-RTS)
- Number of lapses with RTs > 400ms ($PVT_{1/2}$-LAP)

The indexes "$_{1/2}$" refer to the time of measurement (PVT_1 stands for an evening measurement and PVT_2 for the morning measurement). In all calculations of the PVT scores responses within the first minute were skipped in order to compensate for potential effects of adaptation and repetition.

Subjective Fatigue Checkcard (FAT). With the $FAT_{1/2}$ (Samn & Perelli, 1982) the subjective level of fatigue was assessed subsequent to the PVT on day 1 and prior to the tests on day 2.

Performance Tests. All subjects were participants in the first stage of ab-initio selection for ATCos at the DLR (Eissfeldt, 1998). From this data set four typical aptitude tests were chosen to validate the PVT scores:

- Mental Concentration Test (MCT): The MCT involves a combination of different cognitive functions such as visual search, working memory, decoding speed, and simple arithmetic under time pressure.
- Visual Perceptual Speed (VPS): The VPS measures the ability to quickly grasp certain details of visually presented information under limited time.
- Vigilance Test (VIG): The VIG measures sustained attention for monotonous visual and auditory stimuli over a longer time period.
- Choice Reaction Time (CRT): The CRT measures mean reaction times to complex visual stimuli.

Personality. In addition, three performance related personality scales were included in the analysis. The three scales were chosen from DLR's personality questionnaire TSS (Goeters et al., 1993):

- Achievement Motivation (ACH): being ambitious, hardworking, competing
- Rigidity (RIG): being orderly, correctly, punctual, and conscientious
- Vitality (VIT): being active in sports, setting store by fitness, robust

Procedure

The selection tests were administered on two consecutive days. Day 1 has eight hours of performance testing (8am to 5pm) plus one hour lunch break. On day 2 the personality test was administered. 207 subjects completed the PVT and the FAT twice: First, between 5 and 6pm after the regular selection tests had been completed on day 1. Second, at about 9am in the morning of day 2 prior to the personality questionnaire. In this paper, the scores of PVT and FAT are indexed with the time of measurement (1 = evening of day 1; 2 = morning of day 2).

Results

Test-retest correlations were calculated for the PVT and FAT to illustrate the stability of the different scores from the evening after the selection tests to the next morning after one full night of rest (see Table 1). PVT RT measures seem to be more stable than lapses and the subjective fatigue score.

Table 1.
*Test-retest correlations for PVT and FAT from day 1 to day 2 (N = 207). (*p < .05; **p < .01).*

Test	Score	r_{tt}
PVT		
	MRT	.53**
	RTF	.61**
	RTS	.46**
	LAP	.42**
FAT		
	Total	.18**

In order to assess the effect of time of day plus being fatigued after eight hours of testing, scores of PVT and FAT were analysed by t-tests for dependent samples. All effects indicate a significant increase of alertness from evening of day 1 evening to morning of day 2 (see Table 2). Effect sizes are largest for the Fatigue Checkcard followed by PVT lapses.

Table 2.
*Effects of time of day on PVT and FAT scores. Results of t-tests with two-tailed significance levels (*p < .05; **p < .01, N = 207). Effect sizes d were calculated according to Morris & DeShon (2002).*

Test	Score	Day 1	Day 2	d	T
PVT					
	MRT	248.5	244.5	0.18	2.54*
	RTF	204.3	201.3	0.17	2.92**
	RTS	306.3	299.4	0.22	2.95**
	LAP	4.5	3.4	0.31	4.09**
FAT					
	Total	11.2	6.9	1.64	16.18**

In Table 3 the intercorrelations of PVT and FAT scores with the performance tests of the ATCo selection are shown. Correlations are higher for day 1 than for day 2 and again PVT lapses and subjective fatigue show the lowest values.

Table 3.
*Intercorrelations of PVT and FAT scores with performance tests of ATCo selection. N = 247 for day 1 and N = 207 for day 2. (*p < .05; **p < .01).*

Test		MCT	VPS	VIG	CRT	$FAT_{1/2}$
PVT	Score:					
	MRT_1	-.22**	-.19**	-.29**	-.27**	.20**
	RTF_1	-.15*	-.13*	-.20**	-.19**	.14*
	RTS_1	-.24**	-.20**	-.34**	-.28**	.21**
	LAP_1	-.11	-.20**	-.23**	-.20**	.24**
	MRT_2	-.14*	-.16*	-.18**	-.28**	.06
	RTF_2	-.13	-.09	-.06	-.19**	.09
	RTS_2	-.04	-.16*	-.15*	-.21**	.11
	LAP_2	.02	-.11	-.11	-.16*	.00
FAT						
	$Total_1$	-.04	-.03	-.06	-.08	
	$Total_2$	-.12	-.02	.00	.09	

In Table 4 the intercorrelations of PVT and FAT scores with the TSS personality scales of the ATCo selection are shown. It is remarkable that the TSS-personality scale Vitality correlates significantly and consistently with all PVT measures except the lapses. Other correlations of PVT and FAT with TSS-scales (including Extraversion) are negligible.

Table 4.
*Intercorrelations of PVT and FAT scores with personality scales of ATCo selection. N = 207. (*p < .05; **p < .01).*

Test		ACH	RIG	VIT
PVT	Score:			
	MRT_1	.06	.05	-.28**
	RTF_1	.01	.04	-.25**
	RTS_1	.07	.09	-.28**
	LAP_1	.03	.04	-.11
	MRT_2	.04	-.01	-.18**
	RTF_2	.01	.02	-.21**
	RTS_2	.03	.04	-.22**
	LAP_2	-.02	-.15*	-.11
FAT				
	$Total_1$	.06	-.04	-.16*
	$Total_2$	-.04	-.10	.03

To examine whether PVT performance variations across time of day are confounded with trait-related interindividual differences in aptitudes, we correlated the regression residuals for the PVT and FAT from day 2 on day 1 with the performance tests and personality scales of the selection test battery. The changes of FAT due to time of day were uncorrelated to these tests. However, changes in PVT mean reaction times are slightly correlated with the Complex Reaction Time test (-.16**). Changes in PVT lapses are slightly correlated with the TSS-scale rigidity (-.20**). Subjects with higher scores in these two selection tests seem to be slightly more resilient to effects of test fatigue and time of day.

Conclusions

The stability of the PVT from one day to the next is moderate to high (r_{tt} = .42 to .61) and similar to what is reported in the literature (Roach, Dawson & Lamond, 2006). Nevertheless, PVT performance levels showed significant decrements in the evening after subjects had been tested for eight hours compared to the performance level after a full night rest period. The calculated statistics indicate that the RT-scores are more stable and less changeable than the lapses.

The PVT has convergent validities with the Subjective Fatigue Checkcard (r = .20) and with aptitude tests of Vigilance (r = .20 to .34) and Choice Reaction Times (r = .19 to .28). The PVT is also related to the personality scale Vitality (r = .11 to .28). Convergent validities are higher for the evening scores than for the morning scores, which might have been caused by less interindividual variance in the morning.

Finally, it was examined whether performance differences between evening and morning scores of the PVT are in interaction with aptitudes and personality trait measures. Small but significant interactions could be identified for psychomotor speed (correlation with CRT of r = -.16) and for the personality scale Rigidity (r = -.20).

In conclusion, it was found that the PVT scores have both trait and state variance. According to the findings the state variance is larger for the lapses while the RT measures seem to contain more trait variance. Therefore, the lapses are the better indicators of stressor effects on performance in within- and between-subjects designs. However, even the sensitivity of the lapses for such effects can vary for subjects with different levels of Rigidity. Individual vulnerability to effects of stress on performance (in this study test fatigue combined with time of day) seems to be higher for subjects with lower rigidity and with slower psychomotor speed. Therefore, according to the findings presented here, PVT scores cannot be considered as being completely free of aptitude differences.

On the other hand the consistent correlation patterns of the PVT RT measures with typical selection tests have shown that the PVT scores could contribute predictive variance to other tests of attention and concentration within the context of air-traffic controller or pilot selection. Whether a shorter five-minute version of the PVT (e.g. Roach et al., 2006) can still deliver equivalent results for these aptitudes seems worth to be examined.

References

Basner, M. & Dinges, D.F. (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep*, *34(5),* 581-91.

Dinges D.F., Powell J.W. (1985). Microcomputer analysis of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments & Computers*, *17*, 652–655.

Dinges, D.F., Mollicone, D. & Basner, M. (2012). *Psychomotor Vigilance Self Test on the International Space Station (Reaction_Self_Test).* Experiment/Payload Overview (25th July 2012). Retrieved from http://www.nasa.gov/mission_pages/station/research/experiments/Reaction_Self_Test.html.

Dorrian J, Rogers NL, Dinges DF. (2005) Psychomotor vigilance performance: a neurocognitive assay sensitive to sleep loss. In: Kushida C (Ed.), *Sleep deprivation: clinical issues, pharmacology and sleep loss effects (pp. 39–70).* Marcel Dekker, Inc., New York, N.Y..

Eissfeldt, H. (1998) Exemplary Selection Systems - The selection of air traffic controllers. In K.M. Goeters (Ed.) a*viation psychology: A science and a profession (73-80)*. Ashgate, Aldershot, UK.

Elmenhorst D. , Elmenhorst E.-M., Luks N., Maass H., Mueller E.W., Vejvoda M., Wenzel J., Samel A. (2009). Performance impairment during four days partial sleep deprivation compared with the acute effects of alcohol and hypoxia. *Sleep Medicine*, *10*, 189–197.

Goeters, K.-M. und Timmermann, B. und Maschke, P. (1993) The construction of personality questionnaires for selection of aviation personnel. *The International Journal of Aviation Psychology*, *3*, 123-141.

Loh, S., Lamond, N., Dorrian, J., Roach, G. & Dawson, D. (2004). The validity of the psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods, Instruments & Computers*, *36*, 339-346.

Morris, S.B. & DeShon, R.P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105-125.

Roach, G.D., Dawson, D. & Lamond, N. (2006). Can a shorter psychomotor vigilance task be used as a reasonable substitute for the ten-minute psychomotor vigilance task*? Chronobiology International, 23(6),* 1379–1387.

Samn, S.W., Perelli, L.P., 1982. *Estimating aircrew fatigue: a technique with application to airlift operations*. Brooks AFB, USAF School of Aerospace Medicine. Technical report SAM-TR-82-21.

Contact information

Dr. Hans-Juergen Hoermann
German Aerospace Center (DLR)
Institute of Aerospace Medicine
Sportallee 54 a
D-22335 Hamburg, Germany
hans.hoermann@dlr.de

THE UTILITY OF AT-SAT IN HIRING GRADUATES OF AN AIR-TRAFFIC COLLEGIATE TRAINING INITIATIVE PROGRAM

Linda G. Pierce and M. Kathryn Bleckley
FAA Civil Aerospace Medical Institute
Oklahoma City, OK

The FAA recruits applicants for Air Traffic Control Specialist (ATCS) positions from multiple sources. Each hiring source has requirements that applicants must meet for eligibility. These hiring sources include the Air Traffic – Collegiate Training Initiative (AT-CTI), for applicants with specialized education in air traffic control (ATC) and General Public (GP), for applicants with no prior ATC education or experience. Both AT-CTI and GP applicants must take and pass the Air Traffic Selection and Training (AT-SAT), a computerized pre-employment test battery designed to assess an applicant's aptitude for performing the duties of an ATCS. The current research effort compares the selection and training performance of AT-CTI and GP trainees to provide an initial assessment of the utility of AT-SAT as part of the hiring process for AT-CTI graduates.

The Federal Aviation Administration (FAA) hires and trains Air Traffic Control Specialists (ATCSs) to maintain a workforce of approximately 15,000 controllers (FAA, 2012). These ATCSs control air traffic within the National Airspace System at both Terminal and En Route facilities. Terminal facilities include air traffic control towers and Terminal Radar Approach Controls (TRACONs). It is the responsibility of ATCSs within Terminal facilities to organize the flow of air traffic into and out of airports. As air traffic leaves the terminal airspace, the responsibility for control transfers to ATCSs at air route traffic control centers (ARTCC), referred to as En Route facilities.

The Controller Workforce Plan, updated each year, presents the FAA's strategy for hiring, placing, and training controllers to safely meet the demands of air traffic. In 2011, the FAA hired 824 controllers and anticipates selecting more controllers each year through 2020 (FAA, 2012). Multiple hiring sources, with differing criteria are used to solicit applicants. Hiring source and criteria are reflected in the official job announcement. Two controller hiring sources are the General Public (GP) for those without aviation education or experience and the Air Traffic-Collegiate Training Initiative (AT-CTI) program for those with specialized education in aviation.

The AT-CTI program is maintained by the FAA as a collaborative effort with 36 colleges and universities approved to participate in the program. The AT-CTI program produces graduates with a basic understanding of air traffic control. The FAA provides schools in the program with air traffic curriculum, which includes approximately 200 hours of classroom instruction on air traffic control. The schools integrate the FAA-developed coursework into their own 2 or 4-year aviation program. AT-CTI graduates selected for an ATCS position bypass the first five weeks of basic qualification training in air traffic control at the FAA Academy. While they receive no guarantee of employment with the FAA, AT-CTI graduates are considered as a primary hiring source of ATCSs (FAA, 2012).

To be eligible for selection by the FAA as an air traffic controller, AT-CTI graduates and GP applicants must take and pass the Air Traffic Selection and Training (AT-SAT) test battery. AT-SAT is a computerized pre-employment test battery designed to assess an applicant's aptitude for performing the duties of an ATCS. For a detailed description of the development and validation of AT-SAT, see the two volume technical report edited by Ramos, Heil, and Manning (2001a, 2001b). A score of 70 is required to pass AT-SAT. Those applicants who score from 70 to 84.9 are categorized as "Qualified" and those who score from 85 to 100 are categorized as "Well- Qualified."

The FAA engages in an on-going program of research to develop and continually improve strategies to select, train, and place candidates most likely to succeed as air traffic controllers. The purpose of the current research effort is to examine the utility of the AT-SAT test battery as part of the selection process for AT-CTI program graduates. To do so, we examine both the AT-SAT test scores and the training performance of AT-CTI graduates. A GP applicant group is used for comparison. The question examined in this study is whether or not AT-CTI applicants should continue to take AT-SAT as part of the selection process

Method

Sample

We created two samples for our study. The first sample, the applicant sample, included 14,843 applicants for an ATCS position. These applicants took AT-SAT as part of their job application process from April, 2007 through December, 2009. Based on 90% of the applicants reporting gender, the applicant sample was 23.6% female and 76.4% male. Of these applicants, 2,765 were enrolled in and nearing completion or had graduated from an AT-CTI program and 12,078 were applicants from the general public. From the applicant sample, we extracted a second sample, the trainee sample. The trainee sample, included those applicants selected for an ATCS training position by January, 2013. The trainee sample was used to examine training performance of AT-CTI graduates and GP applicants. There were 4,187 ATCS trainees in this second sample: 1,500 AT-CTI graduates and 2,687 GP applicants (18.7 female, 81.3 male).

Measures

We examined AT-SAT scores and AT-SAT category scores (*qualified or well-qualified)* by hiring source (*AT-CTI or GP*) and in relationship to selection decisions and training performance. Selection decisions were o*n board* or no*t on board.* These terms were used because there may be applicants in the database who were selected but have not attended or completed the FAA Academy as of January, 2013 or applicants in the database that may be selected by the FAA after January, 2013. Training performance was captured at the FAA Academy and the trainee's first facility. There were four potential training performance outcomes: *Academy failed or terminated; field training completed successfully; currently in field training; or unsuccessful in field training (facility training failure or transferred to lower level facility).* A fifth outcome was included for those with missing facility training performance data.

Analyses

The first set of analyses used the applicant sample to examine the AT-SAT scores, and category scores, and selection statistics for AT-CTI graduates and GP applicants. The next set of analyses used the trainee sample to compare training performance of AT-CTI and GP trainees at the FAA Academy and the first facility. Throughout the results, we did not report the statistical significance of the differences found in the data. When using datasets with a large number of participants, the likelihood of attaining statistical significance, even with relatively small differences, is high. Therefore, our approach is to describe the data, highlighting differences regardless of statistical significance.

Results

The average AT-SAT score and standard deviation (S.D) for the applicant sample was as follows: AT-CTI (Mean = 87.65, S.D. = 8.23); GP (Mean = 85.69, S.D. = 9.48). On average, the AT-CTI applicants scored 1.96 points higher on AT-SAT than did the GP applicants.

Also examined were the AT-SAT category scores, assigned to AT-CTI graduates and GP applicants. Recall that a minimum score of 70 was needed to pass AT-SAT. The number (N) and percentage (%) of AT-CTI graduates and GP applicants scoring within each of the defined AT-SAT category scores (Not Qualified (< 70), Qualified (70-84.9), and Well-Qualified (85-100)) are shown in Table 1.

Table 1.
AT-SAT Category Scores by Hiring Source

AT-SAT Category Score	AT-CTI N (%)	General Public N (%)
Not Qualified	75 (2.7%)	829 (6.9%)
Qualified	910 (32.9%)	4,318 (35.8%)
Well-Qualified	1,780 (64.4%)	6,931 (57.4%)
Total	2,765 (100%)	12,078 (100%)

Less than three percent (2.7%) of the AT-CTI graduates received an AT-SAT score less than 70, compared to 6.9% of GP applicants. The largest proportion of both AT-CTI graduates and GP applicants were classified as Well-Qualified, with 7% more of the AT-CTI graduates than GP applicants achieving the Well-Qualified category score. A slightly higher percentage of GP applicants achieved a Qualified category score than AT-CTI graduates (see Table 1).

As of January, 2013, 4,187 (28.2%) graduates/applicants who applied from 2007 through 2009 were selected for an ATCS position with the FAA. Of the AT-CTI graduates, 1,500 (54.2%) were selected, while only 2,687 (22.2%) of the GP applicants were selected.

Table 2 shows the AT-SAT score category of those applicants selected for ATCS positions by hiring source. The majority of those selected from both hiring sources scored in the

Well-Qualified score category. A higher proportion of selections were made from the Qualified score category for AT-CTI graduates (25.7%) than GP applicants (5.7%). There were 14 applicants (11 AT-CTI, 3 GP) selected from the Not Qualified score category. It is assumed that these applicants had taken AT-SAT a second time and earned a qualifying score. The dataset used in this study was based on the first administration of AT-SAT. Thus, those scoring in the Not Qualified score category were removed from the dataset. Training performance data will be examined for those who scored as Qualified or Well-Qualified on the first administration of AT-SAT.

Table 2.
Selected Applicants' AT-SAT Score Category by Hiring Source

AT-SAT Score Category	AT-CTI N (%)	General Public N (%)
Qualified	382 (25.7%)	153 (5.7%)
Well-Qualified	1,107 (74.3%)	2,531 (94.3%)
Total	1,489 (100%)	2,684 (100%)

Training Performance

The next set of analyses examined training performance, first at the FAA Academy and then at the first facility assigned. Trainees must pass the FAA Academy before entering training at their first facility. Most trainees were successful at the FAA Academy and were in or had completed training at their first facility (see Table 3). All differences between AT-CTI and GP trainees were less than two percentage points. Slightly more than 50% of both AT-CTI and GP trainees had completed training as of January 2013.

Table 3.
Training Performance by Hiring Source

Performance Category	AT-CTI N (%)	General Public N (%)
Academy		
Failure	98 (6.6%)	191 (7.1%)
Resigned/Declined	15 (1.0%)	29 (1.1%)
First Facility		
Unsuccessful	50 (3.4%)	94 (3.5%)
Successful	753 (50.6%)	1,377 (51.3%)
In Training	373 (25.1%)	618 (23.0%)
Transfer	10 (0.7%)	41 (1.5%)
Passed Academy/Facility Data Missing	190 (12.7%)	334 (12.4%)
Total	1,489 (100%)	2,684 (100%)

We then examined AT-CTI and GP trainees who had completed training either successfully or unsuccessfully. An "Overall Unsuccessful" category was created by combining the FAA Academy failures with those who were unsuccessful at their first facility from Table 3. As shown earlier, a slightly higher proportion of GP trainees were successful than AT-CTI trainees (AT-CTI 50.6%, GP 51.3%). In addition and using the overall unsuccessful measure, we found that a slightly lower proportion of AT-CTI trainees were unsuccessful than GP trainees: AT-CTI 148 (10%); GP 285 (10.6%).

The last step was to examine the relationship among successful and unsuccessful trainees and AT-SAT category scores. For this analysis, we combined the AT-CTI and GP trainee performance data to examine the proportion of successful and overall unsuccessful trainees by AT-SAT score category. As shown in Table 4, a higher proportion of well-qualified trainees succeeded in training (84.1%) than did qualified trainees (76.8%). In addition, qualified trainees (23.2%) were unsuccessful more often than well-qualified trainees (15.9%).

Table 4.
Training Completion by AT-SAT Category Score

	Successful	Overall Unsuccessful	Totals
Qualified	262 (76.8%)	79 (23.2%)	341 (100%)
Well-Qualified	1,868 (84.1%)	354 (15.9%)	2,222 (100%)

Discussion

These data paint a picture of the selection and initial training performance of ATCSs hired from among AT-CTI graduates and GP applicants. These groups are required to take and pass AT-SAT as part of the hiring process. The cost to the FAA for administering AT-SAT is approximately $360 per applicant (L. Waterford, personal communication, February 9, 2012). The question examined in this study is whether or not AT-CTI applicants should continue to take AT-SAT as part of the selection process.

In reviewing the data, we found that the AT-CTI applicants tested from April, 2007 through December, 2009 scored an average of 1.96 points higher on AT-SAT as compared to GP applicants. More than 97% of the AT-CTI graduates and 93% of GP applicants passed AT-SAT and scored as Qualified or Well-Qualified. Given the high pass rate, it is clear that AT-SAT is not eliminating as many applicants, especially those with aviation education, as it did before reweighting (Wise, Tsacoumis, Waugh, Putka, & Hom, 2001). In addition, the majority of both AT-CTI graduates (64.4%) and GP applicants (57.4%) scored in the Well-Qualified score category. However, there was variability in AT-SAT score categories, with more than 30% of all applicants taking AT-SAT scoring in the Qualified score category.

Of those AT-CTI graduates who passed AT-SAT, a majority (54.2%) were selected for an FAA ATCS position by January, 2013. This compared to the selection of only 22.2% of GP applicants who passed AT-SAT. For GP applicants, only 36.5% of those who scored as Well-Qualified on AT-SAT and less than 5% of those scoring as qualified were hired. This finding seems to indicate that the selection panel has a preference for selecting AT-CTI graduates over GP applicants. However, the difference between AT-CTI and GP trainees in training performance at the first facility does not seem to support a preference. The difference in the percent of AT-CTI and GP trainees successfully completing training at the first facility was less than 2 points.

The selection panel also seems to have a preference for hiring those in the Well-Qualified score category, and these data do seem to support that preference. A difference was found between Successful and Overall Unsuccessful completion of training by AT-SAT category score (Qualified or Well-Qualified). Those who had scored as Well-Qualified were successful 84.1% of the time, while 76.8% of those who scored as Qualified were successful; a difference of 7.3%. In addition, qualified trainees (23.2%) were unsuccessful, more than Well-Qualified trainees (15.9%). This difference may support a relationship between AT-SAT category score and training performance. In a recent investigation, Broach et al. (in press) have shown that Well-Qualified ATCS's are more likely to certify as controllers than those scoring as Qualified.

Conclusion and Recommendation

At this time, we conclude that the utility of AT-SAT in hiring AT-CTI graduates may in its ability to categorize them into Qualified and Well-Qualified score categories, information that can be used by the selection panel as a tool to eliminate applicants from consideration. Our recommendation is to continue to use the AT-SAT test battery to hire AT-CTI graduates and, when possible, select graduates who score as Well-Qualified on AT-SAT. We also recommend reanalyzing the dataset when those who had not yet completed their first facility training have completed it, either successfully or unsuccessfully. It would also be useful to examine the data by type of facility assignment (Terminal or En Route). At that time it might be possible to make more definitive statements regarding the training performance of AT-CTI graduates relative to GP applicants and the utility of the AT-SAT test battery in selecting ATCSs.

References

Broach, D., Byrne, C.L., Manning, C.A., Pierce, L.G., McCauley, D., Bleckley, M.K. (in press). *The validity of the Air Traffic Selection and Training (AT-SAT) test battery in operational use.* Washington, DC: FAA Office of Aviation Medicine.

Federal Aviation Administration (2011). A plan for the future: 10-year strategy for the air traffic control workforce 2011-2020. http://www.faa.gov/air_traffic/publications/controller_staffing/media/CWP_2011.pdf. Retrieved 01-23-12.

Federal Aviation Administration (2012). A plan for the future: 10-year strategy for the air traffic control workforce 2012-2021. http://www.faa.gov/air_traffic/publications/controller_staffing/media/CWP_2012.pdf. Retrieved 09-12-12.

Ramos, R.A., Heil, M.C., & Manning, C.A. (2001a). *Documentation of validity for the AT-SAT computerized test battery, Volume I.* (DOT/FAA/AM-01/05). Washington, DC: FAA Office of Aviation Medicine.

Ramos, R.A., Heil, M.C., & Manning, C.A. (2001b). *Documentation of validity for the AT-SAT computerized test battery, Volume II.* (DOT/FAA/AM-01/06). Washington, DC: FAA Office of Aviation Medicine.

Wise, L.L., Tsacoumis, S.T., Waugh, G.W., Putka, D.J., Hom, I. (2001). *Revisions of the AT-SAT* (DTR-01-58). Alexandria, VA: Human Resources Research organization (HumRRO).

Acknowledgements

Research reported in this paper was conducted under the Air Traffic Program Directive/Level of Effort Agreement between the Human Factors Research and Engineering Group (ANG-C1), FAA Headquarters, and the Aerospace Human Factors Research Division (AAM-500) of the FAA Civil Aerospace Medical Institute.

COMPARING THE EFFECTS OF SIMULATED, INTELLIGENT AUDIBLE CHECKLISTS AND ANALOG CHECKLISTS IN SIMULATED FLIGHT

Bryan W. Hilton, Ph.D.
Western Michigan University
Kalamazoo, Michigan
Steve Hard, B.A.
Western Michigan University
Kalamazoo, Michigan

This study evaluated the effect of a simulated, intelligent audible checklist system in a Cessna 172R PC-ATD on checklist errors and omissions. This study was a multiple baseline, time series design, in which participants would fly three to five flights per condition. Each condition used either a 42 item paper or audible checklist. Baseline flights had participants use paper checklist as normal and would graded for correctness. In the intervention phase, at the logical time for a checklist, a red light would illuminate on the dashboard and an audible checklist would engage. An electronic voice would list each item to be completed and the participants were told to say "check" and touch each item as it was completed, intervention flights graded for correctness. During the reversal phase all elements returned to baseline conditions. Baseline condition yielded 22.7% checklist compliance, intervention yielded 97% compliance, and reversal yielded 34% compliance.

Airplane checklists are used during different segments of flight to sequence specific, critical tasks and aircraft adjustments that correspond to specific environmental demands (Degani & Wiener, 1990). These checklists include items for each segment of a flight such as *before take off*, *climb, cruise, descent, before landing,* and *after landing.* Each checklist has specific tasks for each segment, as well as points that work as continued checks on the airplane's configuration in flight. The complexity of these checklists cannot be overstated. For example, on some checklists, the "before engine start" sub-section has 76 items for the first flight of the day, and 37 items for subsequent flight segments (Degani & Wiener). Though many checklists designed by different aircraft companies have similar items, very few are identical. Even different model aircraft made by the same manufacturer typically have different checklists that pertain to different options for those aircraft. Even though these differences exist, checklists have become the main strategy to standardize pilot performance and increase flight deck safety (Rantz, Dickinson, Sinclair, & Van Houten, 2009.) Thus, it is not surprising that many aviation experts have addressed their importance and design, as well as the practices and policies that surround their use (Adamski & Stahl, 1997; Degani, 1992, 2002; Degani & Wiener 1990; Federal Aviation Administration [FAA], 1995, 2000; Gross 1995; Turner, 2001; Rantz and Van Houten, 2011; Rantz, Dickinson, Sinclair, and Van Houten, 2009). Even so, the incorrect use of flight checklists is still often cited as the probable cause or a contributing factor to a large number of accidents (Degani, 1992, 2002; Degani & Wiener; Diez, Boehm-Davis, & Holt, 2003; Turner, 2001). Similarly, many investigations by the National Transportation Safety Board (NTSB) have revealed that the aircraft were not properly configured for flight, which usually results from improper checklist use (NTSB, 1969, 1975, 1982, 1988a, 1988b, 1989, 1990, 1997).

Studies by Lautmann and Gallimore (1987) and Helmreich, Wilhelm, Klinect, and Merritt (2001) provide more direct evidence of improper checklist use by flight crews (Rantz and Van Houten, 2011). In a study funded by Boeing, Lautmann and Gallimore ran a survey of twelve airlines and compiled the data, which showed that errors involved with using the checklist contributed to a substantial occurrence of accidents and incidents. Helmreich et al. conducted a series of studies sponsored by the National Aeronautics and Space Administration to identify the particular errors flight crews commit. Crews were observed while flying. Errors were recorded using the Line Oriented Safety Audit (LOSA) developed by Helmreich and his colleagues (Helmreich, Klinect, Wilhelm, & Jones, 1999; Helmreich et al., 2001). Between 1997 and 1998, LOSAs were conducted at three airlines with 184 flight crews on 314 flight segments (Helmreich et al., 2001). In this study the possible errors were broken down into five categories. Checklist errors fell into the category of "Rule-Compliance" errors. The category of Rule-compliance errors had the highest frequency of errors at 54% of all errors recorded. Checklist errors accounted for the highest number of errors in that category. A similar study using LOSAs between 2002 and 2006 indicate that procedural errors account for about half of all observable errors with checklist errors the most common type within that category (Merritt and Klinect, 2006).

Rantz, Dickinson, Sinclair, and Van Houten (2009) said,

> Despite widespread recognition that checklist errors occurred relatively frequently and were major contributing factors to many crashes, the design of checklists "escaped the scrutiny of the human factors profession" until the 1990s (Degani & Wiener, 1993, p. 28). Degani and Wiener (1990, 1993) observed flight crews while flying, interviewed flight crews from seven major U.S. airlines, and analyzed how the design of checklists contributed to aircraft crashes and incidents that were reported in three aviation databases. Their analytic guidelines became the industry standard (Patterson, Render, & Ebright, 2002).
>
> Although Degani and Wiener (1990) did not pursue the behavioral factors that influence checklist use, they recognized their importance, indicating that safety culture issues erlated to support of misuse or nonuse of checklists were a core problem that led some pilots to misuse the checklist or not use it at all. They also noted that the promotion of a positive attitude toward the use of the checklist procedure was an important element that was often overlooked. Regardless, an extensive search of the aviation checklist literature did not reveal any studies that have examined whether behavioral interventions could increase the appropriate use of flight checklists (p. 498).

Degani and Wiener (1990) mention the creation and description of a computerized device which gives audio prompts of checklist items but they listed no published studies that have compared the use of the audio checklist with any other type of checklist. A review of the literature reveals no comparison studies or behavioral studies of the efficacy of an audio checklist in aviation. Despite the lack of studies there have been several audio checklist devices patented for the use in aviation applications. One device was patented by Harshaw, Burkey, Doell and Keith in1990 and another by O'Rourke in 2001.

Palmer and Degani (1991) indicate that computerized devices that sense completion for the pilot increased checklist errors. In this study, pilots would tend to trust the technology each time they, (a) detected all the mis-configured items when using the paper checklist, (b) detected some of the manual-sensed items displayed by the electronic checklist and (c) did not detect any of the mis-configured item when using the automatic-sense electronic checklist. In the current study, we looked to emulate the function of the paper checklist, only we have removed the need for the physical list during the intervention and workload associated with finding and reading the paper list.

In organizational studies in non-aviation settings (i.e., manufacturing, hotels, banks, offices, retail establishments, and restaurants), checklists have been employed as part of package interventions to improve a diverse array of performances. These performances include: Cleaning and housekeeping tasks (Altus, Welsh, & Miller, 1991; Anderson, Crowell, Hantula, & Siroky, 1988; Anderson, Crowell, Sponsel, Clarke, & Brence, 1982); office tasks (Bacon, Fulton, & Malott, 1982); banquet set-up times (LaFleur & Hyten, 1995); machine set-up time, (Wittkopp, Rowan, & Poling, 1990); metal yield (Moses, Stahelski, & Knapp, 2000); end-of-shift closing tasks (Austin, Weatherly, & Gravina, 2005); staff-client contact time (Porterfield, Evans, & Blunden, 1985); and customer service (Crowell, Anderson, Abel, & Sergio, 1988). Interestingly, none of the studies monitored whether or not employees actually used the checklists. This may be because the checklists were used to inform employees what they were supposed to be doing and thus were considered to be a necessary part of the intervention, but were not viewed as the important motivating variables. The other independent variables that were implemented along with the checklists (i.e., feedback, goals, and rewards) were viewed as the important variables improving compliance. In all of these studies, the checklists were used as an independent variable, not a dependent variable. We were not able to find any study that examined how to increase the use of a checklist in this literature base."

Despite the lack of research in the aviation field there have been studies done in other fields that indicate that the use of audio prompting checklists can increase performance in a variety of situations. Davies, Stock, and Wehmeyer (2002) increased the performance and completion of vocational activities by individuals labeled with mental retardation by the use of a palm top computer which annunciated an audible checklist where items needed to be checked off after they were completed. In this study the use of the palm computer reduced the number of errors per activity from 2.25 to 0.75.

The current study will examine whether digital audio, intelligent checklist system can increase the accuracy of pilots checklist performance compared to a standard paper checklist.

Method

Setting

The experimental setting was the simulation lab in Wood Hall at WMU's main campus. The lab was set up in two adjacent rooms, the simulator room and the control room. The simulated cockpit, cameras, audio system and projector were located in the simulator room. The computers and systems used to control and monitor the flight were located in the adjacent control room.

Participants

Participants for this study were sophomore and juniors enrolled in the WMU College of Aviation Flight Program. Each participant had a single engine instrument rating. Each participant had a minimum of 100 total flight hours up to 272 flight hours, including simulation time with an average of 177 total flight hours and a range of 172 hours. All participants were male and were selected by their willingness to participate and the ability pass the introductory flight session. No female pilots volunteered for this study.

Experimental Task

A personal computer – aviation-training device (PC-ATD) emulating the Cessna 172R was used during the baseline, intervention and reversal phases of this experiment. The flight pattern that participants flew was divided into six segments: (a) pre-takeoff (18 items), (b) after takeoff (2 items), (c) cruise (5 items), (d) arrival (8 items), (e) pre-landing (6 items), and (f) after landing (3 items). The flight pattern was approximately 30 minutes in duration to complete. To realistically simulate an actual flight pattern and insure that it was flown in a consistent manner across trials and participants, the experimenter provided typical air traffic control instructions throughout the flight pattern. These instructions were transmitted using a commercially available intercom system. The speaker was placed next to the PC-ATD and the experimenter, who was in an adjacent area, used the push-to-talk feature on the monitor to transmit the air traffic control instructions. The pilots were instructed to communicate with ATC (experimenter) by speaking aloud. The experimenter, without the participant needing to press any buttons or take any additional steps, could hear all sounds from the simulator room.

During all flights, participants had access to either a paper or simulated audible flight checklist (Independent Variable) and were instructed to use either in the way they would normally use a checklist. Both the paper and simulated audible checklist contained 42 items divided into sections that correspond to each of the six flight segments identified above. The flight checklist was a modified version of the checklist standard to the Cessna 172R. Certain segments on the standard checklist (starting engine, before taxiing, and taxiing) were eliminated to reduce the overall time required to conduct each flight trial. The standard paper checklist was provided to the participant with approach plates. The simulated audible checklist would be initialed by configuration of the plane in-flight- simulating an intelligent system design that monitored flight characteristics such as altitude, flight duration, speed, and vector.

Current paper checklists are currently used in two ways. The first is to follow a challenge-response called a "do list". The pilot must read an individual item and perform the required operation or check before proceeding to the next item on the checklist. The second method is the "flow check". Pilots use memorized flow patterns to check items within a checklist segment. Only after the flow pattern is complete will the pilot return to the checklist to confirm each item is complete. This study provided an audible checklist that replaced both the paper "do list" and memorized "flow list."

It should be noted that the provision of the simulated, intelligent audible checklist system also theoretically emulates a two person flight crew if one of the crew were to provide "challenge and response" to the other pilot in flight.

Method of Data Collection

A multiple baseline across participants design with a reversal to test for maintenance was employed in this study. The design was used to compare simulated audible checklists and standard paper checklists.

The checklist compliance behavior of each participant was remotely monitored via high definition *EZWatchPro* brand digital cameras and software and scored using the checklist observation form (Appendix D). The experimenter/observer would occupy an area adjacent to the PC-ATD . The three digital cameras had built in microphones which allowed the observer to see and hear both the verbal and nonverbal responses required to complete checklist items on the computer monitor and speakers located in the observer's area. Secondarily, a *Sansa* digital audio recorder was also placed next to the PC-ATD to capture audio that may have been too faint for the camera microphones. One camera was mounted above the projection screen approximately 48 inches in front of the participant to capture hand and arm movements. One other was positioned on a wall 5 feet behind the participant to observe the participant's interaction with the flight panel. The third camera was mounted 24 inches to the side rear of the participant to also capture hand and arm movements from a different vantage point. In addition to the cameras, the experimenter was able to see the same screen that each participant saw on a monitor in the separate adjacent "control" room. All flights were recorded and stored digitally for the purposes of conducting inter-observer agreement.

Independent Variable

The independent variable in this study was the presence of either a paper (analog) checklist or a simulated, intelligent audible checklist. The simulated, intelligent audible checklist system was designed by the experimenter and assembled from commercially available components. The system simulated an intelligent computer system that monitored aircraft configuration and automatically started a digital voice that would list the checklist items appropriate for that segment of flight. The trained observers scored accuracy of checklist use. There were three phases (baseline, intervention, and reversal) with a minimum of three to four trials per phase.

Dependent Variables

The main dependent variable consisted of the number of checklist items completed correctly per flight. There were a total of 42 checklist items that could have been counted as correct or incorrect. An item was marked correct if the item was done in the proper order and at the correct time in flight. An item would be marked incorrect if the item was done out of order, done at the wrong time, generally incorrect (performing a task improperly such as setting radios to an incorrect frequency), or omitted. If the participant did a segment of the checklist at an incorrect time, but then also re-did the list at the correct time, the items would be counted as correct. Each item on the checklist was worth one point except for the "pre-flight" checklist. That checklist is different than the other segments in that it must be done twice. The checklist must be done first as a "flow list" from memory and then as confirmatory "do-list" item by item from the actual checklist (not from memory). Because of that difference, this section was graded in half points. The participant would have to complete an item twice to receive the full point. If an item were only done once, the participant would receive ½ a point.

Baseline

Participants were asked to fly one of the four simulated flight patterns (Jackson, MI Airport, Battle Creek, MI Airport, Lansing, MI Airport, or Kalamazoo, MI Airport) three to five times randomly selected during baseline phase. Each flight was a pattern around a different airport than the one they had just previously flown. Four flights were flown per two-hour session under instrument conditions (cloud bottoms at 600 feet and visual sight distance less than three nautical miles. Participants were instructed to use the paper checklist as they would in any normal flight and were also instructed to touch and audibly announce each item as it was completed.

Intervention

The participants were asked to fly one of the four listed flight patterns in instrument conditions with four randomly selected flights per two-hour session. When the time that the participant would traditionally begin using the paper checklist, a red light would illuminate on the dash board; and directly following the illumination of the light, a digital voice would begin to list the checklist items specific to that segment of flight. The simulated audible checklist was a digitized checklist set to replay a digital voice for each item on the list selected by the experimenter. Before the pilot went into the intervention phase, but after the pilot had finished the baseline phase, the pilot was given an explanation of how the intelligent, digital audible checklist would function. Then, each participant was told to announce "check" after the digital audible system listed each item and the item was completed. Once initiated, the system would continue with a brief pause between items, but would not stop unless instructed by the participant. If the items were not completed by the participant and the audible checklist had finished the red light would stay on until the pilot restarted the checklist by saying "restart." The instructional script read to each participant may be found in Appendix F. After they said "check" the next item on the checklist was presented until the checklist was complete at which time the red indicator light on the control panel of the plane would shut off. The participant was further instructed to say "pause" if they needed to pause, and "resume" when they were ready to begin again.

Reversal

During the reversal phase each participant was told that there would no longer be an audible checklist and that experimenter would return the paper checklist to the cockpit. During the reversal phase, each participant was also told to use the checklist as he would in any normal flight and to touch and announce each item as it was completed. The simulated flight flown during the reversal phase was identical to the preceding two conditions. The instructional script read to each participant before a reversal flight was the same script as used in Baseline.

Analysis of Data

For each participant, the number of checklist items completed correctly and incorrectly (omissions included as incorrect) were charted for each session. Changes across the phases were visually analyzed.

Inter-Observer Agreement (IOA)

Inter-observer agreement was calculated on 80% of the total flights across all conditions and participants. For flights that were graded for IOA, two of the trained experimenters would independently grade each flight. The IOA for this experiment was 98.42%.

Results

After the audible checklist was introduced the checklist compliance behavior went from 22.7 % correct in baseline to 97% correct during the intervention. The reversal phase showed a decrease in correct use to 34%, which was somewhat higher than the level obtained during the paper checklist baseline. Each participant showed a marked increase in proper checklist use following the introduction of the audible checklist, but Participant 3 had the widest range of scores from intervention to either baseline or reversal. Much of the reason for the wide range of scores for Participant 3 relates to the fact that while in baseline, the pilot completely omitted setting the navigation system. That omission made the pilot unable to find the airport on two flights despite Air Traffic Control (ATC) rerouting his flight to re-intercept the localizer on approach twice per flight. In reversal, Participant 3 failed to contact the respective airport tower or change radio frequencies to enable contact to the tower on two of the three flights. Those omissions were in direct conflict with instruction from Approach Control to contact the tower and change frequency. The third flight of reversal for Participant 3 once again had a failure to set navigation resulting in multiple missed approaches of the airport.

Acknowledgments

This research was used as Bryan W. Hilton's doctoral dissertation. The authors would like to thank Ron Van Houten, Ph.D., William Rantz, Ph.D., David Fuhrmann, and Laura Hilton.

References

Adamski, A., & Stahl, A. (1997). Principles of design and display for aviation technical messages. *Flight Safety Digest*, *16*(1), 1-29.

Altus, D., Welsh, T., & Miller, K. (1991). A technology for program maintenance: Programming key researcher behaviors in a student housing cooperative. *Journal of Applied Behavior Analysis*, *24*, 667-675.

Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings (1985-1998). *Journal of Organizational Behavior Management, 21*(1), 3-29.

Anderson, C. D., Crowell, C. R., Hantula, D. A., & 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., Clarke, M., & Brence, J. (1982). Behavior management in the public accommodations industry: A three-project demonstration. *Journal of Organizational Behavior Management, 4*(1/2), 33-66.

Austin, J., Weatherly, N. L., Gravina, N. E. (2005). Using task clarification, graphic feedback, and verbal feedback to increase closing-task completion in a privately owned restaurant. *Journal of Applied Behavior Analysis, 38*, 117-120.

Bacon, D. L., Fulton, B. J., & Malott, R. W. (1982). Improving staff performance through the use of task checklists. *Journal of Organizational Behavior Management, 4*(3/4), 17-25.

Balcazar, F. E., Hopkins, B. L., & Suarez, Y. (1985-86). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management, 7*(3/4), 65-89.

Boorman, D. (2001). Today's electronic checklists reduce likelihood of crew errors and help prevent mishaps. *ICAO Journal, 1*, 17-21.

Crowell, C. R., Anderson, D. C., Abel, D. M., & Sergio, J. P. (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.

Davies, D.K., Stock, S.E., & Wehmeyer, M.L. (2002) Enhancing Independent Task Performance for Individuals with Mental Retardation Through the Use of a Handheld Self-directed Visual and Audio Prompting System. *Education and Training in Mental Retardation and Developmental Disabilities,* 37(2), 209-218.

Degani, A. (1992). *On the typography of flight-deck documentation* (NASA Contractor Rep. 177605). Moffett Field, CA: NASA Ames Research Center.

Degani, A. (2002, April). *Pilot error in the 90s: Still alive and kicking*. Paper presented at the meeting of the Flight Safety Foundation of the National Business Aviation Association, Cincinnati, Ohio.

Degani, A., & Wiener, E. L. (1990). *Human factors of flight-deck checklists: The normal checklist* (NASA Contractor Rep. 177549). Moffett Field, CA: NASA Ames Research Center.

Degani, A., & Wiener, E. L. (1993). *Cockpit checklists: Concepts, design, and use. Human Factors, 35*(2), 28-43.

Degani, A., & Wiener, E. L. (1994). *On the design of flight-deck procedures* (NASA Contractor Rep. 177642). Moffett Field, CA: NASA Ames Research Center.

Diez, M., Boehm-Davis, D., & Holt, R. (2003). Checklist performance on the commercial flight-deck. *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 323-328). Columbus, OH: The Ohio State University.

Federal Aviation Administration. (1995). *Human performance considerations in the use and design of aircraft checklists*. Washington, DC: Author.

Federal Aviation Administration. (1996). *Advisory circular 120-64: Operational use and modification of electronic checklists*. Washington, DC: Author.

Federal Aviation Administration. (2000). *Advisory circular 120-71: Standard operating procedures for flight deck crewmembers*. Washington, DC: Author.

Gross, R. (1995). Studies suggest methods for optimizing checklist design and crew performance. *Flight Safety Digest*, *14*(5), 1-10.

Harshaw, Robert C., Ronald S. Burkey, James T. Doell, Dennis G. Keith. (1990). Computerized Checklist System (Heads Up Technologies, Inc.). *U.S. Patent No. 4970683*. Washington, DC: U.S. Patent and Trademark Office.

Helmreich, R. L. (in press). Culture, threat, and error: Assessing system safety. *Proceedings of the Royal Aeronautical Society Conference*. London: The Royal Aeronautical Society.

Helmreich, R. L., Klinect, J. R., Wilhelm, J. A., & Jones, S. G. (1999). *The Line/LOS Error Checklist, Version 6.0: A checklist for human factors skills assessment, a log for off-normal events, and a worksheet for cockpit crew error management* (Tech. Rep. No. 99-01). Austin, TX: University of Texas, Human Factors Research Project.

Helmreich, R. L., Wilhelm, J. A., Klinect, J. R., & Merritt, A. C. (2001). Culture, error, and crew resource management. In E. Salas, C. A. Bowers, & E. Edens (Eds.), *Improving teamwork in organizations* (pp. 305-331). Hillsdale, NJ: Erlbaum.

LaFleur, T., & Hyten, C. (1995). Improving the quality of hotel banquet staff performance. *Journal of Organizational Behavior Management, 15*(1/2), 69-93.

Lautmann, L., & Gallimore, P. (1987). Control of the crew-caused accident: Results of a 12-operator survey. *Boeing Airliner,* 1-6. Seattle: Boeing Commercial Airplane Company.

Merritt, Ashleigh and Klinect, James.(2006). Defensive Flying for Pilots: An Introduction to Threat and Error Management. The University of Texas Human Factors Research Project, The LOSA Collabrative.

Moses, T., Stahelski, A., & Knapp, G. (2000). Effects of attribute control charts on organizational performance. *Journal of Organizational Behavior Management, 20*(1), 69-90.

Mosier, K. L., Palmer, E. A., & Degani, A. (1992). Electronic Checklists: Implications for decision making. In Proceeds of the Human Factors Society 36th Annual Meeting, Santa Monica, CA: Human Factors and Ergonomics Society.

National Transportation Safety Board. (1969). *Aircraft accident report: Pan American World Airways, Inc., Boeing 707-321C, N799PA, Elmendorf Air Force Base, Anchorage, Alaska, December 26, 1968* (Report No. NTSB/AAR-69/08). Washington, DC: Author.

National Transportation Safety Board. (1975). *Aircraft accident report: Northwest Airlines, Inc., Boeing 727-25, 264US, Near Thiells, New York, December 1, 1974* (Report No. NTSB/AAR-75/13). Washington, DC: Author.

National Transportation Safety Board. (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* (Report No. NTSB/AAR-82/08). Washington, DC: Author.

National Transportation Safety Board. (1988a). *Aircraft accident report: BAe-3101 N331CY. New Orleans International Airport. Kenner, Louisiana* (Report No. NTSB/AAR-88/06). Washington, DC: Author.

National Transportation Safety Board. (1988b). *Aircraft accident report: Northwest Airlines, Inc., McDonnell Douglas DC-9-82, N312RC, Detroit Metropolitan Wayne County Airport, Romulus, Michigan, August 16, 1987* (Report No. NTSB/AAR-88/05). Washington, DC: Author.

National Transportation Safety Board. (1989). *Aircraft accident report: Boeing 727-232, N473DA. Dallas-Fort Worth International Airport, Texas* (Report No. NTSB/AAR-89/04). Washington, DC: Author.

National Transportation Safety Board. (1990). *Aircraft accident report: USAir Inc., Boeing 737-400, LaGuardia Airport, Flushing, New York, September 20, 1989* (Report No. NTSB/AAR-90/03). Washington, DC: Author.

National Transportation Safety Board. (1997). *Aircraft accident report: Wheels up landing, Continental Airlines Flight 1943, Douglas DC-9 N10556, Houston, Texas, February 19, 1996* (Report No. NTSB/AAR-97/01). Washington, DC: Author.

Noyes, J. M., Starr, A. F. (2007). A comparison of speech input and touch screen for executing checklists in an avionics application. *International Journal of Aviation Psychology, 11*(3). 299-314.

O'Rourke, James D. (2001). Checklist Device. U.S. Patent Application Publication No. US 2001/0030611 A1. Washington, DC: U.S. Patent and Trademark Office.

Palmer, E., Degani, A. (1991). Electronic checklists: Evaluation of two levels of automation. *Proceedings of the 6th International Symposium on Aviation Psychology* (pp. 178-183). Columbus, OH: The Ohio State University.

Patterson, E. S., Render, M. L., & Ebright, P. R. (2002). Repeating human performance themes in five health care adverse events. *Proceedings of the 46th meeting of the Human Factors and Ergonomic Society* (pp. 1418-1422). Santa Monica, CA: Human Factors and Ergonomic Society.

Porterfield, J., Evans, G., & Blunden, R. (1985). Involving families and staff in service improvement. *Journal of Organizational Behavior Management, 7*(1/2), 117-133.

Rantz, W. G., Dickinson, A., Sinclair, G., Van Houten, R. (2009). The Effect of Feedback on Accuracy of Checklist Completion During Instrument Flight Training. *Journal of Applied Behavior Analysis, 42(3), 497-509.*

Rantz, W.G. & Van Houten, R. (2011) A Feedback Intervention to Increase Digital and Paper Checklist Performance in Technically Advanced Aircraft Simulation. *Journal of Applied Behavior Analysis, 44(1), 144-150.*

Turner, T. (2001). *Controlling pilot error: Checklists and compliance*. New York: McGraw-Hill.

Wilk, L. A., & Redmon, W. K. (1998). The effects of feedback and goal setting on the productivity and satisfaction of university admissions staff. *Journal of Organizational Behavior Management, 18*(1), 45-68.

Wittkopp, C. J., Rowan, J. F., & 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.

FLIGHT DECK AUDIO DISPLAYS: STRIKING THE RIGHT TONE IN FUTURE DESIGNS

Mitchell L. Serber
The MITRE Corporation
McLean, Virginia - USA

The 2009 accident involving Air France Flight 447 illustrates the gap between visual and auditory display design and implementation in commercial aircraft. By examining the successful evolution of Enhanced Ground Proximity Warning System (EGPWS) and Traffic Alert and Collision Avoidance System (TCAS) to incorporate both audio and visual alerting and resolution guidance to flight crews, a contrast is drawn with similar aids available in advance of aircraft upset episodes. A universal approach for future design should incorporate harmonized video and audio displays providing optimized alerting and guidance along with the latest in simulator training designed to emphasize aircraft state awareness degradations before a loss of control incident emerges. By promoting continued dialogue among the aviation community, it is hoped that the existing gap among research, design, and operations may be narrowed, resulting in flight deck audio displays optimized to support effective pilot performance.

This paper explores issues of flight deck audio display design using the backdrop of the 2009 accident involving Air France Flight 447 to illustrate the continuing opportunities to refine and improve this important sensory aid for new generation commercial aircraft. Simpson (2007) laments on early audio display designs: "research literature is replete with examples from earlier times of arbitrary assignment of sounds to cockpit alerting meanings." A recent study surveying the commercial pilot community on auditory alert characteristics indicates loud and persistent alerts may result in adverse unintended consequences, yet remain widespread in modern commercial aircraft design (Peryer, Noyes, Pleydell-Pearce & Lieven, 2005). Aside from cutting-edge military applications, these norms have become firmly established and are indicative of disconnect between the aviation human factors and manufacturer design engineer's communities, ultimately contributing tragic links to the accident chain of many ill-fated flights. By promoting continued dialogue among the aviation community, it is hoped that the existing gaps among research, design, and operations may be narrowed, resulting in flight deck audio displays optimized to support effective pilot performance.

Air France 447

In the early morning hours of June 1, 2009, Air France Flight 447 began to experience a chain of events that quickly resulted in the Airbus A-330-203 entering a deep stall and rapid descent, crashing into the sea near the TASIL reporting point, in international waters of the Atlantic Ocean (BEA, 2011). The aircraft had been cruising at Flight Level 350 when it entered a cloud layer associated with nearby precipitation indicated by onboard radar and subsequently experienced inconsistency between measured airspeed indicators, most likely due to an obstruction of the pitot probes in an ice crystal environment (BEA). Consequently, the invalid airspeed displays resulted in an automatic disconnect of the autopilot along with the associated "cavalry charge" auditory cue (BEA). This disconnect alert was followed four seconds later by the "Stall Stall" audio warning in response to manually entered aft control inputs which caused the aircraft pitch to momentarily exceed the stall warning threshold limits for the given situation (BEA). This first stall warning was a short-lived transient cue that was triggered in advance of an actual stall situation. Later in the accident sequence, as the aircraft actually entered a deep stall condition, the "Stall Stall" audio speech warning and associated "cricket" tonal warning were heard on the Cockpit Voice Recorder (CVR) for a continuous period of 54 seconds (BEA).

Audio Display Design

Drivers

The need to offload the visual system of humans interacting with machines provided the initial stimulus for the development of audio systems to deliver warning messages (Cooper, 1977; DuRoss, 1978). Another advantage sound enjoys over vision is a typically faster processing time, a characteristic often resulting in sound being used effectively to transmit critical information (Petocz, Keller & Stevens, 2008). Unlike the visual system, in order to perceive audio signals, there is no requirement on the user to be positioned in a particular way to receive the signal

(Petocz, Keller & Stevens, 2008). Thus, responding to detected anomalies or system state changes in routing task monitoring may be ideal applications for audio display alerting systems.

Design Philosophy

The laboratory-to-development approach outlined by Patterson (1982) featured purpose-built auditory icons designed as attention getting sounds. Early lab test results were used to formulate a range of parameters based on pulse rate, pitch and dissonance as measured against the perceived urgency among the test subject (Simpson, 2007). Later experiments would compare alert systems where pilots received spoken alerts with preceding attention-getting sounds vs. the same spoken alerts with no attenson preceding the spoken alert (Simpson, 2007). Other researchers experimented by grouping auditory signals by speech, abstract sounds and auditory icons (i.e., environmental sounds with pre-existing associations) (Petocz, Keller & Stevens, 2008). This research established the importance of the association of the strength of the auditory sound with the psychoacoustic knowledge of the subject, the signal referent relationship (Petocz, Keller & Stevens, 2008).

Applied Audio Display – Classic Use Case

Eastern 401

Eastern Airlines Flight 401, operating a Lockheed L-1011 aircraft, crashed 18 miles from Miami, Florida while the flight crew attempted to ascertain whether the nose wheel was in the down and locked position for landing (NTSB, 1973). On the initial approach the crew did not observe an illuminated nose gear down light following the selection of the landing handle to the down position and asked Air Traffic Control (ATC) for vectors to go troubleshoot the situation away from the airport (NTSB). Although the primary cause of this accident was determined to be poor crew coordination, which was manifest by the fact that none of the flight crew was flying the aircraft or actively monitoring its performance, the author believes audio display design played a contributing role to the accident, which claimed 101 lives (NTSB).

Autopilot disconnect. When the Eastern L-1011 crew executed a missed approach because of the suspected nose gear malfunction, they proceeded to climb the aircraft to 2,000 feet, where upon the captain instructed the first officer who was the flying pilot to engage the autopilot (NTSB). Unbeknownst to the crew, the altitude hold feature of the autopilot was disengaged following a suspected control input by the captain as he turned around in his seat to conduct a conversation (NTSB). According to the NTSB report, the altitude disengage feature would not trigger an accompanying “CMD DISC” warning on the captain or first officer annunciator panel and there was no audio warning of an autopilot disengagement available.

Altitude alert. When the aircraft was assigned to maintain the altitude of 2,000 feet Mean Sea Level, that value was entered into the altitude alert window (NTSB). Following the undetected disengagement of the altitude hold feature of the autopilot, the aircraft began a subtle decent towards the ground which was not initially noticed amid the commotion on the flight deck. As the aircraft descended through 1,750 feet the altitude alert system produced an audio warning and the NTSB report noted “a half-second C-chord, which indicated deviation of +/- 250 feet from selected altitude sounded in the cockpit. No crewmember commented on the c-chord. No pitch change to correct for the loss of altitude was recorded.” Therefore, the audio warning was not processed or acted upon by any of the crewmembers in a manner that was recorded and analyzed.

Radar altimeter. The final example of an audio warning in the Eastern 401 accident was issued by the radar altimeter, which was still set for the initial approach sequence to Miami International. As the aircraft neared the ground, ATC noted the flight had departed 2,000 feet and asked the flight how things were going (NTSB). When the crew informed ATC that they wanted to return to the airport, the subsequent clearance triggered the crew to enter a new heading into the auto flight system, whereupon they quickly noted something was amiss with their altitude. Shortly after the captain exclaimed “Hey, what’s happening here” the first of six radio altimeter warning “Beeps” began followed by the sound of impact (NTSB). This audio warning was not designed for this type of situation and by the time this audio was perceived, it was too late to recover the aircraft.

Lessons Learned

While Eastern 401 may be viewed by some in the human factors field as a case study of dysfunctional Crew Resource Management (CRM), the author believes it equally seminal to those researchers studying audio display designs in early commercial jet aircraft. This accident demonstrates the importance of a balanced design approach to facilitate appropriate problem solving under stressful conditions utilizing proper alert intensity, duration and urgency conveyance. Many developments in audio display design have been the result of difficult lessons revealed by this and other related accident investigations.

Enhanced Ground Proximity Warning System (EGPWS)

In a summary of the evolution of terrain avoidance warning systems, Barry C. Breen (1995) of Allied Signal Electronics & Avionics Systems details that the leading cause of worldwide aviation fatalities into the 1990's comes from controlled flight into terrain (CFIT) accidents, where there was nothing remarkably wrong with the aircraft as was the case for Eastern Flight 401. Responding to the high rate of worldwide CFIT accidents, the Commercial Aviation Safety Team (CAST), a joint collaboration of the Federal Aviation Administration (FAA) and industry, successfully developed both technology and training safety enhancements (SEs) designed to mitigate this leading cause of accidents (Angers, 2009). The ground proximity warning system (GPWS) received enhancements including a digital terrain elevation database (DTED) and advanced look-ahead algorithms, which reduced false warnings and provided additional time to execute avoidance maneuvers (Breen). The results of this multi-faceted approach show a significant decline in worldwide CFIT accidents (Boeing Aircraft Company, 2011).

Although the EGPWS information is primarily displayed to the flight crew on the electronic flight information system (EFIS) in the form of the terrain avoidance display (TAD) that appears when the system is manually selected or when a look-ahead warning is triggered, an audio display is also associated with EGPWS. The audio warnings include voice-synthesized warnings such as "Terrain –Terrain" and "Pull-Up Pull-Up" interspersed between a "Whoop Whoop" generated tone (Breen, 1995). The most noteworthy concept of this approach is that the speech portion of the audio display contains two important aspects of interest for the flight crew. First, it informs them of the nature of the threat (terrain) and secondly, it provides them with a corrective action (pull-up). The implementation of EGPWS into the commercial airline fleet, combined with proper crew training, has largely eliminated CFIT accidents in the United States, as FAA's Deputy Associate Administrator for Aviation Safety Peggy Gilligan (2007) reported to Congress in 2007, the last CFIT accident in the U.S. occurred 29 years ago.

Traffic Alert and Collision Avoidance System (TCAS)

Another evolutionary safety enhancement encouraged by the CAST, is the traffic alert and collision avoidance system (TCAS) (Angers, 2009). TCAS is an airborne system developed by the FAA that operates independently from the ground-based Air Traffic Control (ATC) system and contains both visual and audio displays (Searight, 2010). Designed to increase cockpit awareness of proximate aircraft and to serve as a "last line of defense" for the prevention of mid-air collisions, the latest versions of TCAS include speech generated warning and guidance output similar to EGPWS (Searight, 2010). The evolution of TCAS has continued with the refinement of resolution advisories that include 3-D audio representation of the traffic conflict, depicted with reference an out-of the-cockpit view (Begault, Wenzel & Lathrop, 1997). An outgrowth of contemporary military systems designed to optimize flight deck control/display interfaces, the 3-D audio feature may soon make its way onto the commercial flight deck of the future (Taylor, et al., 2001).

Survey Says

Shock and Awe

Ask any commercial pilot, and you will be certain to get an opinion about the nature of audio alerting in transport category aircraft. When Peryer, Noyes, Pleydell-Pearce and Lieven (2005) conducted their survey, they noted the association between high intensity audio sounds and both undesirable physiological and psychological effects. Highest on the pilot survey of criticisms of auditory alerts is the unnecessarily extreme volume (Peryer, et al.). Their study also cited Aviation Safety Reporting System (ASRS) database reports declaring loud alerts as an adverse factor in flight safety, sometimes resulting in the immobilization of crews subjected to the startle factor of the alert (Peryer, et al.).

Make it Stop

Pilots interviewed for a study on approaches to improving auditory and visual alerting in aircraft expressed that certain audio alerts had durations that were inappropriately prolonged (Ulfvengren, Martensson & Singer, 2002). In some cases, the lack of ability to cancel the alert following acknowledgment was seen as contributing to perceptual and cognitive overload, undermining the ability of the pilot to understand and evaluate two keys for successful problem solving. Peryer, et al. (2005), citing previous work notes:

> The auditory modality is very sensitive to change, which makes it ideal for warning presentation; however, if the alert continues, attention, even at a rudimentary level, is still devoted to processing the auditory signal and is diverted away from processing task-related information (Banbury et al., 2001; Wickens & Hollands, 2000).

While these surveys were conducted with subjects who fly advanced-automation aircraft representing all of the major airframe manufacturers, it is noteworthy that pilot criticism of auditory alerting found throughout these surveys has not substantively changed over the last 20 years (Peryer, et al.). Patterson (1982) notes that design engineers have "flooded the flightdeck" with piercing sounds which are not only disliked, but more importantly disrupt thought and inhibit communication during critical events. This comes in spite of the dramatic gains to the visual displays on the same aircraft (Peryer, et al.).

Aircraft State Awareness

A Complex Problem

While much of this paper has discussed events involving aircraft avoidance of objects, EGPWS avoiding terrain and TCAS avoiding aircraft, a more confounding problem in flight deck design appears to be audio display of aircraft performance state related to in-flight upset incidents. It is significant that EGPWS and TCAS are fundamentally geo-spatial problems to solve, being less dependent on specific aircraft aerodynamic performance characteristics. Additionally, there is often more time to accurately predict and present alerts and solution guidance both visual and aural, for EGPWS and TCAS. Aircraft state awareness involves a considerably greater dynamic set of variables including meteorological aspects, which along with a specific airframe's aerodynamic properties and operational differences may contribute to aircraft upset incidents like a high altitude stalls and ice induced roll upsets.

Don't Get Upset

According to Sunjoo Advani, chairperson of the International Committee for Training in Extended Envelopes, organized by the Royal Aeronautical Society, when an aircraft is upset it is not only outside the normal flight envelope, but also in an agitated condition (Warwick, 2011). Pilots in these types of situations need to quickly and accurately correlate the stimuli presented, including aircraft generated audio alerts in order to affect a recovery. Aircraft behavior in an upset may seem more startling and perplexing to flight crews whose only previous exposure is likely to have been in a training simulator (Warwick, 2011). The audio displays associated with these events need to support and complement existing visual cues, not conflict and confound an already task-saturated flight crew.

Closing the Visual-Auditory Display Gap

Ulfvengren, Martensson and Singer (2002) describe each of the major design philosophies employed by aircraft manufacturers, in leading up to their discussion on improving aircraft audio designs. The "Sort and Guide" model they state, was inspired by the A320 model of alert logic prioritization in which immediate action items are presented one at a time, in priority order, with clear fault information presented along with guidance for recovery (Ulfvengren, Martensson & Singer, 2002). Examination of Figure 1 below will show the many auditory alerts generated by the A330 aircraft flown by Air France Flight 447 (A330 flight deck, 1999). It should be noted that with respect to aircraft state awareness, the synthetic voice alerts generated provide no direct guidance for recovery, unlike those previously described for EGPWS and TCAS alerts. Stanton and Edworthy (1999) suggest development of new audio designs should focus on the linkage between the sound and the potential for remedial action, referents should guide action required of the pilot. Air France Flight 447, and similar in-flight upset accidents, show a need for further research and development of audio displays to fully actualize the concept of 'Sort and Guide".

WARNING SIGNAL	CONDITION	DURATION	SILENCING
CONTINUOUS REPETITIVE CHIMIE	RED WARNINGS	PERMANENT	Depress* MASTER WARN lt
SINGLE CHIME	AMBER CAUTION	1/2 sec.	
CAVALRY CHARGE	A/P DISCONNECTION BY TAKE OVER pb	1.5 sec	Second push on TAKE OVER pb
	A/P DISCONNECTION DUE TO FAILURE	PERMANENT	Depress MASTER WARN lt or TAKE OVER pb
CLICK	LANDING CAPABILITY CHANGE	1/2 sec (3 pulses)	
CRICKET + "STALL" message (synthetic voice)	STALL	PERMANENT	NIL
INTERMITTENT BUZZER	SELCAL CALL	PERMANENT	Depress RESET key on ACP
BUZZER	CABIN CALL	3s	NIL
	EMER CABIN CALL	3s REPEATED 3 TIMES	NIL
	MECH CALL	As long as outside pb pressed	NIL
	ACARS ◁ CALL or ALERT	PERMANENT	Message reading on MCDU or Depress MASTER CAUT
C CHORD	ALTITUDE ALERT	1.5 sec or PERMANENT	new ALTITUDE selection or depress MASTER WARN pb
AUTO CALL OUT (synthetic voice)	HEIGHT ANNOUNCEMENT BELOW 400 FT	PERMANENT	NIL
GROUND PROXIMITY WARNING (synthetic voice)	UNSAFE TERRAIN IN CLEARANCE FORESEEN	PERMANENT	NIL
"WINDSHEAR" (synthetic voice)	WINDSHEAR	REPEATED 3 TIMES	NIL
"PRIORITY LEFT" "PRIORITY RIGHT" (synthetic voice)	A/PTAKE OVER pb	1 sec	NIL
"RETARD" (synthetic voice)	THRUST LEVER NOT IN IDLE POSITION FOR LANDING	PERMANENT	THRUST LEVER
TCAS ◁ (synthetic voice)	TRAFFIC OR POTENTIAL COLLISION	PERMANENT	NIL

Figure 1. A-330 Electronic instrument system – ECAM audible warning definitions. Synthetic voice provided guidance for EGPWS and TCAS highlighted in green.

Discussion

The tragedy of Air France Flight 447 should be a call to the aviation industry to evaluate how to best leverage the use of automation in concert with advanced upset training for pilots. A universal approach for future design should incorporate a combination of harmonized video and audio displays with the latest in simulator training. This would provide optimized alerting and guidance in a training environment designed to emphasize aircraft state awareness degradations before a loss of control incident emerges, resulting in aircraft control being continually maintained. Many challenges remain to be solved to enable audio display design to complement visual alerts, and not issue contradictory or confound information in an emerging upset scenario. With the near elimination of CFIT as the number one cause of accidents in the US, it is time to redouble our efforts and apply the lessons learned from CFIT to other areas of aviation safety. With solid research, development, validation and implementation, it is possible to have a similar reduction on the Loss of Control (LOC) accident rate as has been realized with Controlled Flight Into Terrain (CFIT) accidents. Striking the right tone in future audio display designs will help eliminate one link in the accident chain.

References

A330 flight deck. (1999). *A330 flight deck and systems briefing for pilots*. Blagnac Cedex: AIRBUS. Retrieved from http://www.smartcockpit.com

Boeing Aircraft Company. (2011). *Statistical summary of commercial jet airplane accidents worldwide operation 1959-2010*. [Online]. Available: http://www.boeing.com

Angers, S. W. A. (2009). Safety in numbers: Industry team recognized for improving aviation safety. *Boeing Commercial Airplanes: Boeing Frontiers*, Retrieved from http://www.boeing.com

BEA. (2011). *On the accident on 1st june 2009 to the airbus a330-203 registered f-gzcp operated by air france flight af 447 rio de janeiro- paris* . Retrieved from Bureau d website: http://www.bea.aero

Begault, D. R., Wenzel, E. M., & Lathrop, W. B. (1997). *Augmented tcas advisories using a 3-d audio guidance system*. Paper presented at the Ninth International Symposium on Aviation Psychology Ninth international symposium on aviation psychology, the ohio state university, Columbus, Ohio. Retrieved from http://human-factors.arc.nasa.gov

Breen, B. C. (1999). Controlled flight into terrain and the enhanced ground proximity warning system. *IEEE AES Systems Magazine*, (January), 19-24. Retrieved from http://ieeexplore.ieee.org

Cooper, G. E. (1977). *A survey of the status and philosophies relating to cockpit warning systems* (Contract Report Number NAS2-9117)NASA Ames.

DuRoss, S. H. (1978). *Civilian aircraft warning systems: A survey of pilot opinion within british airways* (Technical Report Number 78056). United Kingdom: Royal Aircraft Establishment.

Gilligan, P. (2007). *Statement of peggy gilligan, deputy associate administrator for aviation safety: Before the committee on transportation and infrastructure, subcommittee on aviation on the most wanted list of the national transportation safety board* . Retrieved from Federal Aviation Administration website: http://www.faa.gov

NTSB (1973). *Aircraft accident report eastern airlines, inc., l-1011, n310ea, miami, florida, december 29, 1972* (NTSB-AAR-73-14). Retrieved from National Transportation Safety Board website: http://NTSB.gov

Peryer, G., Noyes, J., Pleydell-Pearce, K., & Lieven, N. (2005). Auditory alert characteristics: A survey of pilot views. *The International Journal of Aviation Psychology*, *15*(3), 233-250. Retrieved from http://www.interruptions.net

Patterson, R. D. Civil Aviation Authority, (1982). *Guidelines for auditory warnings on civil aircraft* (CAA Paper Number 82017). London: Civil Aviation Authority.

Petocz, A., Keller, P. E., & Stevens, C. J. (2008). Auditory warnings, signal-referent relations, and natural indicators: Re-thinking theory and application. *Journal of Experimental Psychology: Applied*, *14*(2), 165-178. Retrieved from http://marcs.uws.edu.au

Searight, J. S. (2010). *Faa tcas home page*. Retrieved from Federal Aviation Administration website: http://adsb.tc.faa.gov

Simpson, C. A. (2007). *Doing science on auditory display design in the cockpit: Merging laboratory rigor and the aircraft cockpit environment*. In *Proceedings of the 13th International Conference on Auditory Display* (pp. 139-142). Retrieved from http://dev.icad.org

Stanton, N. A., & Edworthy, J. (1999). Auditory warning affordances. In N. Stanton & J. Edworthy (Eds.), *Human factors in auditory warning* (pp. 113-127). Aldershot, England: Ashgate Publishing.

Taylor, R. M., et al., (2001). *Cognitive cockpit engineering: Coupling functional state assessment, task knowledge management and decision support for context sensitive aiding*. Retrieved from Human Systems Information Analysis Center, Department of Defense website: http://131.84.179.51

Ulfvengren, P., Martensson, L., & Singer, G. (2002). *Improving audio and visual alerting in aircraft by means of part-task simulation*. In Clemens Weikert (Ed.), *Human Factors in Aviation Proceedings of a Conference September 26-27, 2002, Lund, Sweden* (pp. 92-100). Retrieved from http://citeseerx.ist.psu.edu

Warwick, G. (2011, November 28). Push for safety: Industry moves to close pilot training gaps exposed by growing loss-of-control accidents. *Aviation Week and Space technology*, *173*(42), 44-46.

Acknowledgements

The author acknowledges MITRE colleague, Valerie Gawron Ph.D. and Mr. Troy Faaborg, Embry-Riddle Aeronautical University Worldwide Campus for their mentoring, inspiration, and encouragement to apply my operational expertise and research interest in sensation and perception issues to modern flight deck design. The views contained in this report do not reflect the views of the MITRE Corporation.

INDIVIDUAL DIFFERENCES IN PERCEPTION AND PERFORMANCE OF ADVANCED NAVIGATION SYSTEMS

Andre Garcia, Jesse Eisert, John Payne, Carryl L. Baldwin
George Mason University
Fairfax, VA

Victor Finomore
Air Force Research Lab
Wright Paterson Air Force Base, Dayton, OH

We examined individual differences in use and preference for tactile route guidance formats. Participants drove a simulated vehicle through counterbalanced pairings of four distinct cities using one of four navigation systems (three tactile and one auditory control). One tactile system used only pulse rate, the second system used only tactor location, and the third used both pulse rate and location to convey guidance instructions. All navigation systems provided both a preliminary and an immediate cue indicating to take the next most immediate turn. Individual differences in sense of direction resulted in different preference ratings without any observed performance differences. The pulse-rate route guidance system was the most commonly preferred system, especially for those with a poor sense of direction. All four systems resulted in equivalent wayfinding performance and support previous literature indicating that tactile guidance systems can effectively support navigation in unfamiliar environments, even for individuals with poor sense of direction.

Tactile technology is becoming more prevalent in various new in-vehicle systems such as lane departure warnings and driver alertness warnings. The Ford Motor Company's "Lane Keeping System" informs the driver they are drifting out of their lane by vibrating the steering wheel. Mercedes Benz also uses a steering wheel vibration to alert the driver to pay attention to the road when the system senses the driver is fatigued or drowsy. For example, in the current Cadillac XTS, General Motors added a feature that vibrates the seat when the driver may be backing up into an object that cannot be seen. Tactile technology has also been used in a variety of settings for navigation purposes.

The sense of touch is generally an under-utilized modality and may be used to relay information to the user in an un-obstructive and minimally invasive way. Merlo, Duley, & Hancock (2010) used a vibrotactile belt at the waist to relay traditional Army hand signals to infantry who may not be within eyesight. Van Erp & Van Veen (2006) demonstrated that a vibrotactile waist belt can be an easy to learn and intuitive route guidance system in waypoint navigation. Garcia, Finomore, Burnett, Baldwin, & Brill (2012) found results consistent with Van Erp & Van Veen (2006) for dismounted soldiers traversing through a virtual environment of a Middle East war zone.

Van Erp & Van Veen (2004) investigated the use of a vibrotactile seat for in-vehicle navigation in normal and high workload conditions. They found that a tactile navigation display can help reduce the workload involved with driving, particularly in high workload settings. Yet few of these studies have investigated the individual differences that may exist with the use of tactile technology for navigation purposes.

Individual differences in navigation strategy based on sense of direction or spatial abilities have been extensively investigated. Garcia, et al. (2012) demonstrated that individuals with a good sense of direction (GSD) based on the Sense of Direction Questionnaire (SDQ) (Kato & Takeuchi, 2003) were significantly faster and more accurate navigators traversing through a virtual environment. Individuals with a good sense of direction are better at maintaining their heading in relation to their cardinal heading, whereas those with a poor sense of direction tend to use a verbal approach to navigation and benefit most from egocentric route guidance instructions (Baldwin & Reagan, 2009). Individuals with a good sense of direction benefit most from allocentric visual based route guidance systems which allow them to build a better, more global cognitive map of their environment (Furukawa, Baldwin, & Carpenter, 2004).

Due to these individual differences in navigation abilities, individuals may differ in the type of route guidance system they prefer. Individuals may subjectively prefer a certain type of navigation display based on the system's presentation modality characteristics and the type of information that is included in the system. This subjective preference may even conflict with the system design that they would perform best with. This experiment was intended to examine this research question. Specifically, we sought to examine whether individuals with different sense of direction abilities would differ in terms of which vibrotactile route guidance system format they most preferred and whether or not those preferences would also be reflected in navigation performance. Based on previous research (Baldwin and Reagan, 2009) we reasoned that individuals with a good sense of direction might be more likely to prefer a tactile system that did not disrupt their use of visuo-spatial working memory resources during route learning. Of the tactile systems examined, the system that uses tactor location to convey guidance information is the most likely to involve visuospatial working memory resources and therefore we predicted it would be the least favored system among individuals with a good sense of direction. It was further predicted that individuals with a good sense of direction would commit fewer turning errors than individuals with a poor sense of direction and that they would have relatively better overall route recall regardless of the navigation format used. Furthermore, it is predicted that the redundant route guidance system would be the most effective at conveying route guidance instructions overall.

Methods

Participants

57 undergraduate participants from George Mason University provided written informed consent and then participated in this experiment. All reported normal or corrected to normal vision and hearing and were recruited from the undergraduate population.

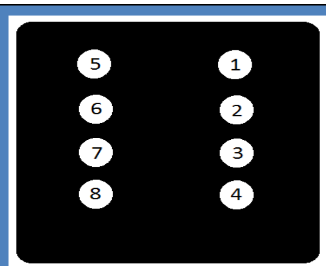
Apparatus

A driving simulator created by RealTime Technologies, Inc. was used for this experiment. The simulator is capable of yaw and pitch motion. The yaw motion allows for 180 degrees of motion, 90 left and 90 right and the pitch motion allows for 1.5 degrees of pitch motion to simulate abrupt acceleration and braking. Virtual/physical rotating motion were decoupled to a .5:1 ratio, meaning that for every 90 degrees of motion in the virtual world, the simulator only turned 45 degrees in the physical world. The simulator features (3) 42” plasma high definition screens that allows for 180 degree forward field of view. The cab was built from a 2002 Ford Taurus and is operated similar to a real car with an automatic transmission.

The simulator is equipped with a 5.1 surround sound speaker system and a vibrotactile seat that contains 8 tactors arranged in 2 rows of 4 C2@ tactors. The vibrotactile seat was custom designed and constructed by Engineering Acoustics, Inc. Three different tactile route guidance systems were designed in addition to the more traditional auditory route guidance system. The first tactile route guidance system, known as the “redundant” system, gives a preliminary route guidance instruction by vibrating the front half of seat in the given direction of the next turn in a slow pulse rate, and a fast pulse rate in the back half of the seat in the given direction of the next turn for the immediate route guidance instruction.

In the “pulse rate” route guidance system, participants were given a preliminary route guidance instruction by vibrating the middle two tactors on the appropriate side at a slow pulse rate and for the immediate route guidance instruction the middle two tactors were activated at a fast pulse rate. In the “location” route guidance system, an intermediate intensity pulse rate was used in the front half of the seat for the preliminary route guidance cue and in the back half of the seat for the immediate route guidance cue.

This allows for systematic evaluation of the effects of pulse rate or location, or the additive combination of the location and pulse rate of a route guidance cue. This also allowed us to examine whether individual differences in route guidance design preferences based on sense of direction exist. The details of each tactile route guidance condition are summarized in Table 1, below. For further details on the tactile seat, see Garcia, Eisert, & Baldwin (2013).



Condition	Preliminary	Immediate
Auditory Equivalent	“Your next turn will be a [direction]”	“Make the next [direction]”
Redundant	Pulse rate 3.69 Tactor 5+6 For Left Turn Tactor 1+2 for Right Turn	Pulse rate 11.93 Tactor 7+8 for Left Turn Tactor 3+4 for Right Turn
Location	Pulse Rate 7.87 Tactors 5 + 6 for Left Turn Tactors 1+2 for Right Turn	Pulse Rate 7.87 Tactors 7+8 for Left Turn Tactors 3+4 for Right Turn
Pulse Rate	Pulse rate 3.69 Tactors 6+7 for Left Turn Tactors 2+3 for Right Turn	Pulse rate 11.93 Tactors 6+7 for Left Turn Tactors 2+3 for Right Turn

Table 1. *Details of each route guidance format condition and type of cue.*

Procedure

After signing an informed consent document, participants completed the Kato & Takeuchi Sense of Direction Questionnaire (SDQ, Kato & Takeuchi, 2003). Next, participants were escorted into the driving simulator and were given a demonstration of the various features of the simulator. Participants were then given a quick tutorial and training session of how the route guidance systems in general function, were informed on the order of the experiment and were shown a few sample images of the task they would be performing throughout the experiment.

Participants drove through four different cities, twice each (one city per route-guidance system) for a total of 8 experimental drives, and a practice drive before each city to help familiarize the participant with each route guidance system. After each drive, participants were asked to retrace each route they drove on a blank map of the city. They were given the starting locations for each drive before beginning the experimental task. Participants were shown three unique landmarks to attend to as they drove each experimental drive; they were asked to indicate on the map their locations after each pair of drives. Next, participants were given a blank compass to indicate where they thought the point of origin was in relation to their egocentric orientation at the end of each drive.

Experimenters recorded how many turning errors were committed by the participants during the drives as well as the type of errors committed. At the end of the experiment, once the participant was able to complete a pair of drives with each type of route guidance system, participants were asked which route guidance system they preferred. Last, participants were debriefed on the true purpose of the experiment and were given the contact information of the PI in case they had any additional questions.

Results

Due to data collection failures, user's route guidance system preference was collected for only 40 subjects. Overall, participants overwhelmingly preferred the "pulse rate" route guidance system (22), followed by the "redundant" route guidance system (12) and the "location" route guidance system (6). For the purposes of maximizing responses, sense of direction grouping for preference data was determined by simply grouping people as to whether they were above or below the sample mean on the SDQ. The results are organized in table 2 and figure 1.

	PSD	GSD
Redundant	4	7
Location	3	3
Pulse rate	12	9

Table 2. *Route Guidance System Preference Split by SOD*

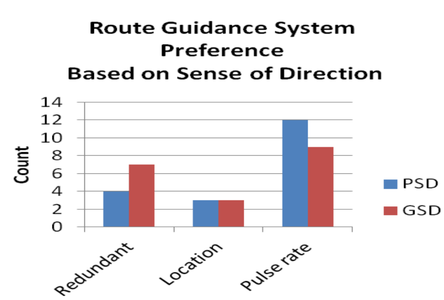


Figure 1. *Route Guidance System Preference Split by SOD*

For the performance data, sense of direction groupings was based on criterion of SSDQ score plus or minus one standard deviation from the mean of a larger sample containing 250 responses. In this grouping, individuals with a good sense of direction had an average score of 2 or below and those with a poor sense of direction had an average of 2.7 or higher. Note that an attempt to use a more stringent criterion would have maximized the potential to observe differences between the groups but at the cost of even further reduction in sample size and statistical power. A 2 (sense of direction X 4 (RGS format) mixed repeated measures ANOVA was used to assess the number of navigational errors (turning errors) committed by participants while using the various route guidance systems. Only 8 individuals met the criteria needed to qualify as a poor sense of direction individual. Overall, there was no statistically significant difference between the 4 route guidance systems in the amount of turning errors made by participants, F (1.3) = 1.24, $p > .05$, as well as no statistically significant interaction, F (1.3) = 2.27, $p > .05$.

Discussion

Tactile route guidance systems have shown great potential in a variety of settings and have begun finding their way into the modern vehicle. The purpose of this experiment was to assess participants' subjective route guidance system interface preference as well as the amount of navigational errors committed.

Overall, the pulse-rate route guidance system was the most preferred vibrotactile navigation system. When analyzing preferences based on sense of direction, the pulse-rate route guidance system was still the most preferred route guidance system. However, individuals with a good sense of direction picked the redundant route guidance system almost as often as the pulse rate route guidance system. We believe this is due to the ability of GSD individuals to understand and benefit from the added spatial information that the redundant route guidance system offers above and beyond the pulse rate RGS. That is, the pulse rate route guidance system distinguishes between a preliminary route guidance cue and an immediate instruction based on a slow or fast pulse rate coming from the same location. The redundant RGS adds the element of spatial location. Not only does it offer the same pulse rate information from the pulse-rate RGS, but the seat also vibrates in the front half (at a slow pulse rate) for a preliminary cue and in the back half on the appropriate side (at a fast pulse rate) for the immediate cue. The added spatial information may have been ignored or found to be an annoyance for those with a PSD, whereas those with a GSD may have experienced the addition of spatial information as a benefit.

There was no significant difference in the amount of turning errors committed between GSD and PSD individuals. There are a few potential explanations for this. Due to the extremely low rate of turning errors, it is a possible that the task was too simple, thus creating a ceiling effect. Furthermore, all four types of route guidance systems were egocentric in nature. This perspective is most intuitive for route guidance purposes, but may not lend itself best to route learning or building a cognitive map of the environment. Additionally, individuals with a poor sense of direction often perform best with interfaces with an egocentric perspective (Baldwin & Reagan, 2009; Garcia et al. 2012). Future investigations should include a condition from a geocentric perspective, similar to Garcia et al. (2012). A limitation of the current investigation is the small sample sizes obtained with our groupings based on the SDQ questionnaire. Small samples sizes resulted in low statistical power and likely contributed to the present non-significant findings. The convenience sample used made it difficult to find enough participants who scored more than one standard deviation above and below the mean on the SDQ. Data collection may continue in the near future to collect more data from individuals who meet the criteria to be classified as a GSD or PSD individual.

Ideally, a multimodal system may be most beneficial in a fully commercial route guidance system. Human factors design principles should be responsibly implemented in commercial multimodal systems so that the navigational cues are perceived as a single gestalt rather than cues being perceived each as a different message for each modality. Additionally, commercial route guidance systems should have the ability to be customizable based on an individual's spatial abilities and preferences.

Acknowledgements

We would like to thank George Mason University's Center of Excellence in Neuroergonomics, Technology, and Cognition (CENTEC) for their continuous support of this project as well as the Air Force Research Lab's Human Performance Wing for their support and collaborative efforts.

References

Baldwin, C. L., & Reagan, I. (2009). Individual Differences in Route-Learning Strategy and Associated Working Memory Resources. *Human Factors, 51*(3), 368-377.

Garcia, A., Finomore, V., Burnett, G., Baldwin, C.L., Brill, C. (2012). Individual Differences in Multimodal Waypoint Navigation. *Proceedings of the 56th Annual Meeting of the Human Factors & Ergonomics Society*. Boston, MA.

Cholewiak, R. W., Brill, J.C., & Schwab, A. (2004). Vibro-tactile localization on the abdomen: Effects of place and space. *Perception and Psychophysics, 66,* 970-987.

Furukawa, H., Baldwin, C. L., & Carpenter, E. M. (2004). Supporting Drivers' Area-Learning Task with Visual Geo-Centered and Auditory Ego-Centered Guidance: Interference or Improved Performance? In D. A. Vincenzi, M. Mouloua & P. A. Hancock (Eds.), *Human Performance, Situation Awareness and Automation: Current Research and Trends, HPSAA II* (pp. 124-129). Daytona Beach, FL

Garcia, A., Eisert, J., Baldwin, C.L. (2013). Comprehension of Vibrotactile Route Guidance Cues. *In Proceedings of 15th International Conference on Human Computer Interaction*. Las Vegas.

Garcia, A., Finomore, V., Burnett, G., Baldwin, C. L., Brill, C. (2012). Individual Differences in Multimodal Waypoint Navigation. In *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*.

Merlo, J.L., Duley, A.R., & Hancock, P.A. (2010). Cross-modal congruency benefits for combined tactile and visual signaling. *American Journal of Psychology, 123* (4), 413-424.

Van Erp, J.B.F., Van Veen, H.A.H.C., Jansen, C., Dobbins, T. (2006). Waypoint Navigation with a Vibrotactile Waist Belt. *ACM Transactions on Applied Perception*. 2(2), 106-117.

Van Erp, J.B.F., Van Veen, H.A.H.C. (2004). Vibrotactile in-vehicle navigation system. *Transportation Research Part F 7*, 247-256.